U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

MESTRADO EM CIÊNCIA DA INFORMAÇÃO

Gestão de Dados de Investigação em Monitorização Ambiental

Patrícia Gabriela Silva Dias

M

2018

UNIDADES ORGÂNICAS ENVOLVIDAS

FACULDADE DE ENGENHARIA
FACULDADE DE LETRAS

PATRÍCIA GABRIELA SILVA DIAS

Gestão de Dados de Investigação em Monitorização Ambiental

Relatório de Projeto realizado no âmbito do Mestrado em Ciência da Informação, orientada pelo Prof. Gabriel David e pelo Prof. Sérgio Nunes

Membros do Júri:

Prof. António Lucas Soares

Faculdade de Engenharia da Universidade do Porto

Prof. Mariana Curado Malta

Instituto Superior de Contabilidade e Administração do Porto

Prof. Gabriel David

Faculdade de Engenharia da Universidade do Porto

Faculdade de Engenharia da Universidade do Porto

Julho de 2018

"Persistence is the shortest path to success"

Charles Chaplin

Agradecimentos

Este espaço é certamente limitado para agradecer a todos as pessoas que ao longo da dissertação e destes cinco anos, me ajudaram, direta ou indiretamente, a cumprir os meus objetivos e a realizar mais uma importante etapa na minha formação.

Ao meu orientador Professor Doutor Gabriel David por me acolher neste projeto, pela partilha de conhecimento, orientação e suporte para a realização desta dissertação.

Ao meu coorientador Professor Doutor Sérgio Nunes pela disponibilidade demonstrada.

A todos os professores que ao longo destes cinco anos contribuíram para que a conclusão desta etapa fosse possível, pelo saber e experiência que nos foi passada.

À Joana pela ajuda incansável, disponibilidade constante, por me incentivar nesta longa caminhada, pelo carinho, paciência e amizade.

A todos os colegas do InfoLab, que me acolheram da melhor forma possível e se mostraram sempre dispostos ajudar, pelos bons momentos passados, em especial ao João Rocha pela ajuda incansável com o Dendro.

Aos investigadores do projeto UrbanSense, que contribuíram para a realização deste projeto, sem os quais não seria possível de concluir.

À minha família com os quais partilharei sempre todas as conquistas, aqueles a quem dedico esta conquista que infelizmente não poderão celebrar junto de mim, mas que eu tenho a certeza que estão sempre a cuidar de mim.

Um especial agradecimento aos meus pais um enorme obrigada por acreditarem em mim, pelo exemplo que me deram, pelos ensinamentos, apoio e dedicação que sempre tiveram comigo e por me proporcionarem a oportunidade de estudar, por me mostrarem que tudo é possível.

Ao Rui, por me apoiar sempre, por estar ao meu lado nestes cinco anos e por me incentivar a fazer sempre mais e melhor.

Aos meus amigos, que desde o início estiveram presentes em todos os momentos que estarão sempre até ao fim e que fizeram destes cinco anos os melhores até então.

À minha colega de casa, companheira de todas as horas, por todos os momentos passados, pelas horas de sono perdidas, pelas conversas e desabafos.

À Cátia, pela amizade, por mesmo longe estar sempre comigo.

Obrigada,

A todos os que me ouvem sem pressa, aos que me abraçam e tiram o peso da vida, aos que me conhecem nos bons e maus momentos e permanecem. Aos que me mostram o lado bom de tudo, a todos os que gostam mesmo de mim!

Resumo

A gestão de dados de investigação tem assumido uma importância crescente no contexto das atividades dos investigadores de muitas áreas. O desenvolvimento das tecnologias utilizadas para produzir, recolher e analisar dados resultou numa maior preocupação com a gestão de dados por parte dos investigadores e entidades envolvidas nos projetos de investigação. Uma vez que são produzidas grandes quantidades de dados, mobilizando elevadas competências, esforço e recursos, a sua simples publicação pode constituir um resultado de investigação, para além da publicação de artigos que os analisem. A publicação dos dados visa também fomentar a sua reutilização por terceiros, potenciando quer investigação independente quer a verificação dos resultados obtidos a partir desses dados.

A presente dissertação estuda de que forma a gestão de dados de investigação pode ser uma mais valia na valorização e reutilização dos dados, sendo que se utiliza como caso de estudo dados ambientais recolhidos por sensores espalhados pela cidade do Porto, resultantes do projeto UrbanSense. Este projeto teve como principal objetivo tornar a cidade do Porto uma Smart City. Uma vez que se trata de uma grande quantidade de dados é importante a sua organização e descrição. Para isso são estudados vários modelos de metadados de onde são selecionados descritores que são incorporados na plataforma colaborativa de gestão de dados—Dendro, com o intuito de depositar os dados no repositório de dados do Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC TEC) para que estes estejam acessíveis e possam ser reutilizados dando origem a novos projetos de investigação.

Palavras-chave: gestão de dados de investigação, curadoria digital, esquemas de metadados, UrbanSense, dados ambientais, smart cities

Abstract

Research data management has become increasingly important in the context of the activities of researchers in many areas. The development of the technologies used to produce, collect and analyze data has resulted in a greater concern with the management of data by researchers and entities involved in research projects. Since large amounts of data are produced, mobilizing high skills, effort and resources, their simple publication can be a result of research, in addition to the publication of articles that analyze them. The publication of the data also aims to encourage their re-use by third parties, enhancing either independent research or verification of the results obtained from these data.

The present dissertation studies how the data management of research can be an asset in the valorization and reuse of data, and it is used as case study environmental data collected by sensors located in the city of Porto, resulting from the UrbanSense project. This project had as main objective to make the city of Porto a Smart City. Since it is a large amount of data, it is important to organize and describe it. In order to do this, several metadata models are studied from which descriptors are selected and incorporated into the collaborative Dendro data management platform, with the purpose of depositing the data in the data repository of the Institute of Systems Engineering and Computers, Technology and Science (INESC TEC) so that these are accessible and can be reused giving rise to new research projects.

Keywords: research data management, digital curation, metadata schemas, UrbanSense, environmental data, smart cities

Sumário

Lista de Figuras

Lista de Tabelas

Siglas e Abreviaturas

Aws	Ontology for Meteorological Sensors
CF	Climate and Forecast
CO	Monóxido de Carbono
DC	Dublin Core
DCC	Digital Curation Centre
DCMI	Dublin Core Metadata Initiative
DCU	Data Collection Units
DDI	Data Documentation Initiative
DTN	Delay Tolerant Network
EU H2020	European Commission Horizon 2020
FEUP	Faculdade de Engenharia da Universidade do Porto
FIESTA-IOT	Federated Interoperable Semantic IoT Testbeds and Applications
FoaF	Friend of a Friend
INESC TEC	Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência
INFOLAB	Information Systems Research Group
IoT	Internet of Things
JISC	Joint Information Systems Committee
NetCDF	Network Common Data Form
NO2	Dióxido de nitrogénio
O&M	Observation and Measurements
O3	Ozono
OBUs	Unidades on-bord
OECD	Organisation for Economic Co-Operation and Development
OGC	Open Geospatial Consortium
OWL	Web Ontology Language
PGD	Plano de Gestão de Dados
RCAAP	Repositório Científico de Acesso Aberto em Portugal
RDF	Resource Description Framwork
RDM	Research Data Management
SSN	Semantic Sensor Network Ontology
TIC	Tecnologias de Informação e Comunicação
V2I	Veículo-infraestrutura

V2V	Veículo a veículo
W3C	World Wide Web Consortium
XML	Extensible Markup Language

1 Introdução

1.1 Contexto e Motivação

O desenvolvimento das tecnologias utilizadas para produzir e recolher dados, resultou numa maior preocupação por parte dos investigadores com o aumento gradual dos mesmos. Este aumento progressivo de dados, apelidado por muitos de dilúvio de dados, resultou na afirmação das práticas de gestão de dados de investigação.

Neste contexto, a motivação para a realização desta dissertação surgiu com base na plataforma UrbanSense, plataforma essa que faz parte do projeto Cidades do Futuro (Future Cities[1]), que objetivou a transformação da cidade do Porto numa Smart City, intitulado de PortoLivingLab.

O PortoLivingLab é uma estrutura desenvolvida no âmbito das Smart Cities que apontam para a melhoria da qualidade de vida dos cidadãos, permitindo uma melhor compreensão dos processos urbanos da cidade do Porto e é composta por três plataformas de monitorização, nomeadamente, o SenseMyCity, o BusNet e o UrbanSense. Estas plataformas atuam em paralelo desde 2013 na recolha de dados essencialmente ambientais, captados através dos vários sensores espalhados pela cidade do Porto. Esses dados consistem por exemplo, em localizações geográficas, qualidade do ar, temperatura, humidade, direção do vento, entre outros.

Os investigadores reconhecem a importância da visualização dos dados. Tendo em conta os benefícios da utilização destes dados, surgiu a oportunidade de analisar e organizar os mesmos para serem posteriormente publicados num repositório de dados de investigação. A publicação dos dados no repositório tem como objetivo torná-los acessíveis a qualquer utilizador, desde os investigadores ao público em geral, como é o caso de possíveis empresas que tencionem utilizar os dados recolhidos para desenvolver serviços para os seus clientes.

Para que os dados sejam proveitosos e resultem na obtenção de informação útil é necessário que os mesmos sejam de qualidade, para isso é necessário avaliar a qualidade dos dados, uma vez que, foi identificado pelos investigadores a existência de conjuntos de dados considerados "inválidos", sendo que os sensores que os recolheram nem sempre estavam calibrados ou em locais propícios.

Atualmente, os dados estão sob a tutela da Faculdade de Engenharia da Universidade do Porto (FEUP) que juntamente com o Instituto de Engenharia de Sistemas e

[1] Projeto *Future Cities*: https://futurecities.up.pt/site/

Computadores, Tecnologia e Ciência – INESC TEC viu, através do projeto de dissertação, uma forma de implementar estratégias de gestão de dados de investigação, nos dados recolhidos pela plataforma UrbanSense.

1.2 Enquadramento do Projeto

1.2.1 Ameaças identificadas

Ao iniciar a realização deste projeto, foi necessário conhecer o projeto "UrbanSense", o que consequentemente, levou à identificação de alguns problemas intrínsecos aos dados que serão analisados ao longo da realização da dissertação.

Inicialmente percebeu-se que os dados não estavam estruturados e ainda que não existia uma distinção clara entre dados brutos e dados processados. O mesmo se verifica na inexistência de um plano de gestão de dados (PGD), que se elaborado antes da recolha dos dados, permitiria que os mesmos estivessem num formato correto, bem organizados e anotados. Os dados existentes são apenas dados brutos o que pode posteriormente resultar numa dificuldade de análise, uma vez que ainda não sofreram qualquer tipo de tratamento.

Outro problema identificado, relacionado em parte com a falta de um PGD corresponde à inexistência de apontamentos sobre os dados, o que pode dificultar a sua compreensão, visto serem dados de uma área que implica conhecimento especializado, mesmo numa fase seguinte de depositar os dados na plataforma de gestão de dados-Dendro. Para contrariar este cenário, seria aconselhável a criação de um PGD, que permite descrever os dados, como se pretende disponibilizá-los e ainda como preservá-los.

Ainda ao nível da compreensão dos dados, outra dificuldade identificada é a linguagem em que os mesmos estão expressos. Um dos objetivos da realização do projeto é que os dados sejam disponibilizados para qualquer pessoa que tenha interesse e isso será feito através da descrição na plataforma de gestão de dados Dendro[2] que permitirá aos utilizadores contextualizarem-se no projeto o que faz com que qualquer pessoa consiga entender os dados e consequentemente utilizá-los.

1.2.2 Objetivos e resultados esperados

Esta dissertação centra-se na gestão de dados de investigação do projeto UrbanSense, e surge com base no potencial de utilização dos dados recolhidos. Os investigadores viram uma oportunidade na partilha de dados com o objetivo da sua reutilização e, por isso, a sua gestão é imprescindível.

[2] Disponível em: http://dendro-stg.inesctec.pt/

Os objetivos do trabalho são diversos (ver Figura 1), e surgem a partir do objetivo central, que consiste na gestão dos dados de investigação, assegurando a qualidade dos dados através análise dos mesmos.

O objetivo central é composto por três outros objetivos, que correspondem às três grandes fases do projeto de investigação, que passam pela produção/ recolha dos dados que neste caso é feita através dos sensores que compõe a plataforma UrbanSense, a fase de organização e qualidade dos dados da plataforma UrbanSense e a última fase que corresponde à disponibilização dos dados com o objetivo de serem reutilizados.

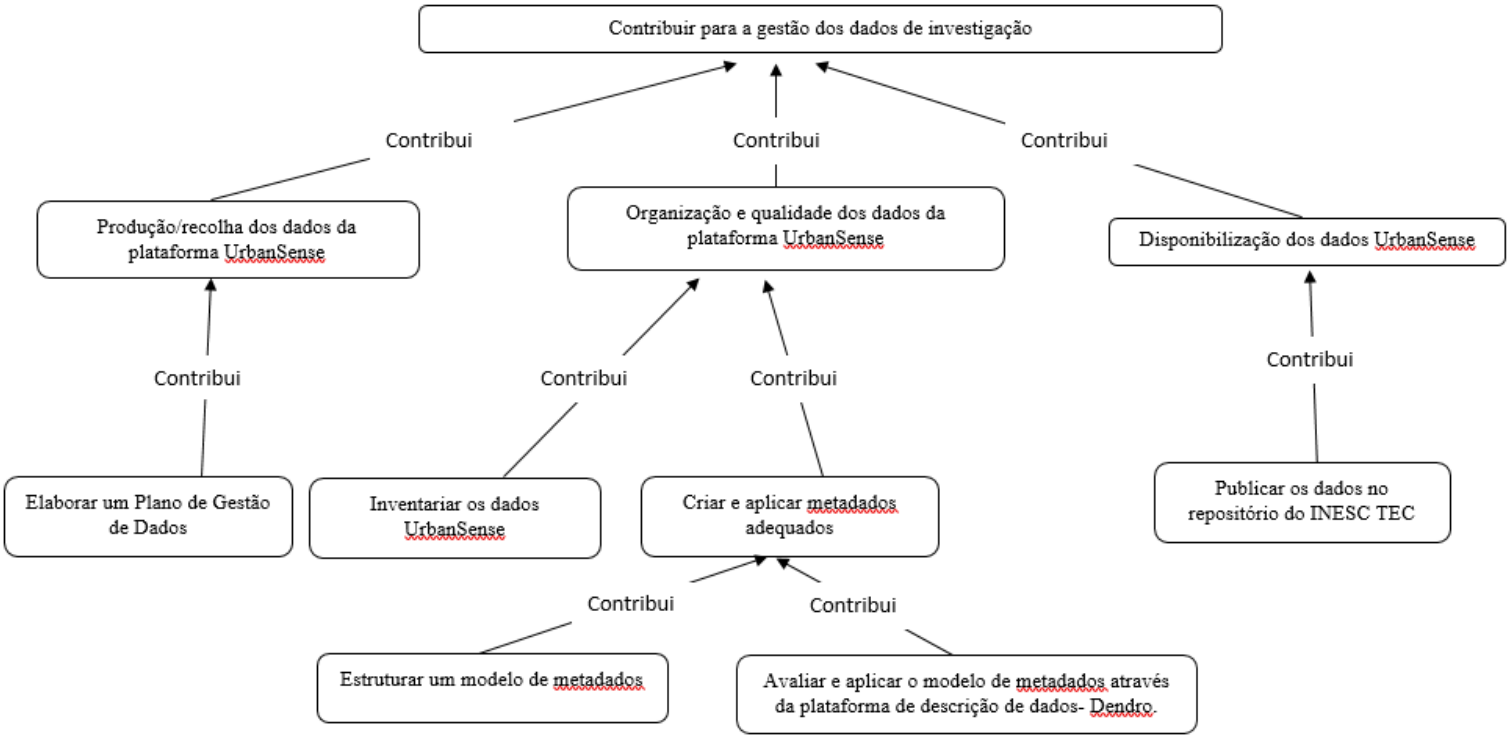


Figura 1 Árvore de objetivos.

Esta dissertação terá o foco ao longo dos três objetivos do segundo nível uma vez que algumas das tarefas que serão realizadas estão envolvidas com as três fases e os objetivos que estarão mais próximos do trabalho prático realizado ao longo da dissertação, são também eles essenciais para todo o processo. A elaboração de um Plano de Gestão de Dados apesar de ser uma tarefa que deveria ser realizada antes da recolha dos dados, irá fazer parte desta dissertação, uma vez que não foi elaborado nenhum PGD e é importante para suportar a preservação dos dados.

As tarefas que envolvem a segunda parte do projeto, que diz respeito à organização e qualidade dos dados da plataforma UrbanSense, começam com a inventariação dos dados UrbanSense, o que permitirá conhecer os dados existentes. Segue-se a criação e aplicação de metadados. Esta fase inicia-se com a estruturação de um modelo de dados, que pode ser

combinada com alguns modelos de metadados já existentes e que podem ser reutilizados. O modelo obtido é em seguida aplicado através da plataforma de descrição, desenvolvida pelo Information Systems Research Group (InfoLab) e que permite gerir os dados, o Dendro.

Na última parte deste projeto, serão disponibilizados os dados UrbanSense, para que estejam acessíveis para outros investigadores, através da publicação dos dados no repositório do INESC TEC. Esta fase é também composta por uma outra tarefa que consiste na avaliação dos resultados, em que será pedido a um investigador que avalie a descrição realizada nos dados UrbanSense.

1.3 Estrutura da dissertação

Esta dissertação está dividida em oito capítulos, mais duas secções destinadas às referências bibliográficas e aos anexos. No primeiro capítulo, destinado à introdução, estão apresentados assuntos como o contexto e motivação, enquadramento do projeto, nomeadamente ameaças identificadas e objetivos e resultados esperados e por último a estrutura definida para a dissertação.

O segundo capítulo, destinado à revisão de literatura, está subdividido em cinco secções principais. Estas secções estão relacionadas com o tema principal, que é a gestão de dados de investigação. Numa primeira parte é abordada a questão dos dados de investigação, na qual é realizada a contextualização do tema. Com base neste mesmo tema foram abordados dois subtemas, o primeiro consiste no ciclo de vida e a gestão dos dados de investigação e o segundo nos Planos de Gestão de Dados (PGD), em que são apresentados os aspetos principais que compõem um PGD. Num segundo tópico, é abordada a curadoria de dados e, no terceiro, questões relacionadas com os metadados. Neste, apresentam-se alguns modelos de metadados de domínios diferentes. A quarta secção é relativa aos formatos de representação de dados. Na última secção, tendo em conta o objetivo principal da curadoria de dados, de garantir o acesso contínuo e ativo aos dados, para que estes possam ser utilizados no futuro por outros investigadores, abordam-se várias questões relacionadas com a preservação de dados de investigação, as implicações que esta tem na gestão de dados e algumas medidas a ter em conta, aquando do arquivo dos dados em repositórios.

No terceiro capítulo são apresentados a abordagem metodológica e os procedimentos que serviram de base para a realização desta dissertação.

O quarto capítulo é dedicado ao projeto UrbanSense, dando-se a conhecer o projeto PortoLivingLab e mais detalhadamente a plataforma UrbanSense.

No quinto capítulo é apresentado o Plano de Gestão de Dados criado para o projeto UrbanSense, uma das primeiras tarefas desta dissertação. Nesta parte do trabalho pretende-se descrever o que será feito com os dados recolhidos no projeto.

No sexto capítulo é apresentada informação sobre os dados da plataforma UrbanSense, incluindo um inventário dos dados, como ponto de partida para a respetiva organização, e a análise das entrevistas aos investigadores do projeto. As últimas secções deste capítulo estão dedicadas ao *workflow* de gestão de dados do projeto UrbanSense, desde a análise e seleção de modelos de metadados à sua aplicação usando a plataforma Dendro, terminando na publicação dos dados no repositório.

O sétimo capítulo destina-se à avaliação dos resultados, em que o foco estará em perceber se o trabalho até então realizado, nomeadamente os descritores utilizados para descrever o projeto, serão suficientes para contextualizar um investigador que não esteja familiarizado com o projeto UrbanSense, mas que tencione reutilizar os dados.

O oitavo capítulo é dedicado às conclusões do trabalho realizado e às perspetivas em termos de trabalho futuro.

Por último encontram-se as referências bibliográficas e os anexos, que contêm os vários materiais desenvolvidos ao longo do trabalho, complementando o que está descrito nos capítulos anteriores.

2 Gestão de Dados de Investigação

No século XXI, os dados produzidos e as tecnologias digitais passaram a ser um tema na agenda no atual contexto científico, estimulado pelos novos métodos e ferramentas. A investigação científica é cada vez mais ativa e tem ganho mais visibilidade na recolha de dados em quase todas as áreas. "Este tipo de ciência já é designado como data-intensive science compreendendo, segundo os seus proponentes, três actividades essenciais: a recolha, a curadoria e a análise de dados" (Rodrigues et al. 2010).

Perante isto, muitas são as questões que se colocam, tanto em relação à produção de dados que tem atingido um volume desmedido, como à diversidade dos dados que resultam das várias atividades de investigação. Por isso necessitam que seja realizada uma descrição que compreenda quer as condições de recolha dos dados quer as condições de acesso aos mesmos.

A questão da gestão de dados de investigação está no centro das preocupações das comunidades científicas sendo a base para a validação dos resultados da investigação. A gestão de dados quando introduzida no planeamento de um projeto torna os resultados mais visíveis o que pode atrair novas parcerias. Eynden, Corti, Woollard, Bishop, & Horton (2011) defendem que a gestão de dados é decisiva para facilitar a partilha de dados e garantir a acessibilidade dos dados a longo prazo e a sua reutilização para futuras Investigações.

Com as constantes necessidades de os utilizadores dependerem de dados digitais, e as dificuldades por parte das organizações em lidar com a produção de grandes volumes de dados, a curadoria de dados é um ponto crucial que interfere nas várias etapas do ciclo de vida, mantendo os dados autênticos e reutilizáveis. A disponibilização dos dados é uma parte integrante deste desafio, quando se pensa neste assunto, que não se pode apenas focar em recolher ou produzir dados, mas sim em produzir, armazenar e disseminar esses dados de forma a garantir a confiança e disponibilidade dos dados para os utilizadores.

Tal como referido acima, a reutilização dos dados é uma questão que contempla a curadoria de dados e também um requisito que leva a que as comunidades científicas sintam necessidade de garantir o acesso e partilha dos dados recolhidos em projetos de investigação, como é o caso do retratado nesta dissertação, para que numa fase posterior os mesmos possam ser reutilizados por pessoas que tenham interesse nos dados.

Quando se fala em disponibilizar e reutilizar os dados, impreterivelmente tem que se falar em preservá-los para que seja possível assegurar-lhes o acesso permanente, questão esta que pode ser vista como um outro desafio, visto que a preservação tem como principal dificuldade a evolução tecnológica.

O ciclo de vida dos dados de investigação, a curadoria de dados, os metadados e a preservação, são questões que se abordam ao longo deste capítulo.

2.1 Dados de investigação: contexto e definição

Com o aumento da relevância dos dados de investigação, a recolha e o tratamento de dados ganharam um estatuto de resultado de investigação de per se, ou enquanto suporte a outros resultados, como artigos. Esta evolução aumentou a exigência sobre os metadados.

A evolução da produção de dados está bem patente no aparecimento de dois novos termos: *big data* e *e-science*. O primeiro remete para conjuntos de dados, em primeiro lugar, de grande volume, mas também de grande dinamismo e grande diversidade (Costa and Cunha 2014; Hey and Trefethen 2003). Mais interessante no contexto desta dissertação é o conceito de e-science. A crescente disponibilidade, na Internet, de repositórios de dados de qualidade permite que equipas de investigação desenvolvam a sua atividade a partir desses repositórios, sem terem elas próprias que produzir os seus dados. É a isto que chama e-science. A reutilização corresponde à redução da fase de recolha de dados e à aceleração do ciclo da investigação. As equipas podem assim concentrar-se, por exemplo, em validar resultados anteriores, em testar hipóteses alternativas para os mesmos fenómenos ou em fazer pesquisa exploratória para identificar novos fenómenos.

O termo e-science é atribuído a John Taylor, diretor geral do Conselho do National e-Science Center no Reino Unido, em 2000 (Costa and Cunha 2014) e é caracterizado pelo uso de sistemas informáticos, pelo trabalho colaborativo e pelos dados de investigação e foi introduzido para designar novas formas de colaboração e investigação multidisciplinar em áreas-chave da ciência e das infraestruturas necessárias para proporcionar novas formas de trabalho (Hey and Trefethen 2003).

Segundo John Taylor, o termo e-science, é definido como:

> "ciência em grande escala que será crescentemente executada através de colaborações globais distribuídas possibilitadas pela Internet. Tipicamente, uma característica de tal empreendimento científico colaborativo é que irá requisitar acesso a grandes coleções de dados, recursos computacionais em grande escala e visualização de alto desempenho para os cientistas" (Gattelli 2015).

Tendo em conta as definições apresentadas por Taylor, conclui-se que o autor dá ênfase à utilização da tecnologia computacional na investigação, o que fez com que o termo

e-science apareça frequentemente na literatura científica e tenha sido alvo de discussões, o que contribui para uma melhor definição do termo.

Há ainda quem defina o termo *e-science* como uma "metodologia de investigação que aumenta as capacidades e os recursos, que estão rapidamente em transformação não só na ciência mas em outras áreas, tornando-se um efeito global" (Digital Curation Centre 2006).

Independentemente das definições existentes, a adoção de novas práticas pela chamada "Sociedade da Informação", levou ao crescimento da preocupação com os dados de investigação, o que fez com estes passassem a ter mais importância já que, segundo Furtado, Príncipe, & Carvalho (2017) os "dados de investigação científica constituem a base para todo e qualquer resultado científico e, consequentemente, para toda a publicação ou output científico, baseada em medições, observações ou pesquisas".

Para Ribeiro, Amorim, Silva, Castro, & Lopes (2016)

> "os dados de investigação são criados e usados em contextos diversos. Podem ser gerados especificamente para um projeto de investigação, como dados de sensores captados numa experiência, ou entrevistas para uma análise. Mas também incluem dados recolhidos sistematicamente para algum fim e que também são usados em investigação, como dados meteorológicos ou os registos de acesso a um serviço de computação".

No seu Guia de Gestão de Dados, Eynden, Corti, Woollard, Bishop, & Horton (2011) explicam que "se os dados de investigação estiverem bem organizados, documentados, preservados e acessíveis e se a precisão e validade dos dados for controlada em todos os momentos, podem-se obter dados de alta qualidade que contribuem para pesquisas eficazes, economizando-se tempo e recursos". Há ainda quem defenda que os dados podem ser vistos como bens públicos e que, por isso, devem estar acessíveis para a sociedade em geral.

Para a OECD (2007) os dados de investigação são definidos como "registos factuais (pontuações numéricas, registos textuais, imagens e sons) utilizados como fontes primárias para a investigação e que são regularmente aceites na comunidade científica como sendo necessários para validar os seus resultados.". A definição dada pela Organisation for Economic Co-Operation and Development (OECD) vai ao encontro da definição dada pela European Comission (2016) no documento de "Guidelines to the Rules on Open Access to

Scientific Publications and Open Access to Research Data in Horizon 2020", onde definem dados de investigação como sendo,

> "informações, em particular factos ou números, recolhidos para serem analisados e considerados como base para fundamentação, discussão ou cálculo. Num contexto de investigação, exemplos de dados incluem estatísticas, resultados de experiências, medições, observações resultantes de trabalho de campo, registos de entrevistas e imagens".

Num documento desenvolvido por Furtado et al. (2017) o Repositório Científico de Acesso Aberto em Portugal (RCAAP) considera dados de investigação "todos e quaisquer dados de investigação que sejam produto direto ou indireto do processo de investigação científica e por isso necessários para a validação de resultados científicos".

Os dados de investigação podem então classificar-se em dois tipos: dados em bruto, que correspondem a dados que não são processados, como por exemplo dados que são recolhidos por sensores e que ainda não sofreram qualquer tipo de tratamento, e dados processados, ou seja, dados que já estiveram no seu estado bruto e que foram manipulados de alguma forma e que são mais facilmente interpretáveis por utilizadores comuns.

Segundo a National Science Board (2005), os dados de investigação podem ainda ser distinguidos segundo as suas origens:

Observacionais - dados registados num exato momento não podendo ser recolhidos novamente e que são arquivados indefinidamente.

Computacionais - resultados da execução de modelos computacionais ou de simulação, não sendo necessária a sua preservação, visto que podem ser reproduzidos.

Experimentais - dados que eventualmente podem ser reproduzidos com precisão e que não necessitam de ser armazenados indefinidamente.

A partilha de dados é uma das formas de validação do método científico, que permite a confirmação dos resultados e o progresso da investigação, com base em resultados prévios, que se encontram agrupados em repositório de preservação dos dados. No entanto o principal objetivo é fomentar o acesso e reutilização dos dados. Assim, para aumentar a partilha e reutilização dos dados, é necessário motivar os investigadores e os curadores de dados para que estes pratiquem a gestão tendo como preocupação todas as fases do ciclo de vida dos dados.

2.1.1 Ciclo de vida e gestão de dados de investigação

Quando nos referimos à gestão de dados temos que ter em atenção em que consiste o ciclo de vida dos dados já que este pode auxiliar os investigadores a compreender o verdadeiro significado da gestão de dados. Para que os investigadores consigam documentar todo o processo, desde como é feita a recolha de dados até como os dados brutos passam a dados processados e garantam que os dados são descritos de forma a serem compreensíveis por outros investigadores e reutilizados, é necessário que seja realizada a gestão de dados ao longo de todo o processo, desde a criação/ recolha à análise de dados.

Perante o volume cada vez maior, a reutilização é o principal objetivo da gestão e preservação de dados de investigação. Desta forma, o RCAAP considera que existem várias formas para os dados de investigação serem reutilizados: os dados primários podem resultar em dados secundários, através da aplicação da seleção e combinação com outros dados, criando-se novos dados de investigação, ou seja, é reiniciado o ciclo de dados. Outra forma é a utilização dos dados para citações em literatura ou estudos científicos, não havendo a criação de novos dados (Furtado, Príncipe, and Carvalho 2017).

Pennock (2007) discute a necessidade de abordar o ciclo de vida de forma a garantir a preservação da informação digital para que a mesma possa ser utilizada por terceiros. O autor no mesmo artigo afirma ainda que apesar da constante evolução tecnológica e das mudanças organizacionais, é possível verificar a proveniência dos dados digitais se se relacionar a gestão de informação com o ciclo de vida. O ciclo de vida assegura a identificação, o planeamento e a implementação correta de todas fases de gestão de dados, o que garante a autenticidade, integridade, confiabilidade e usabilidade dos dados de investigação.

Segundo Furtado et al. (2017), "um ciclo de gestão de dados de investigação procura integrar os conceitos e os momentos-chave no planeamento e implementação de ações relacionadas com a gestão de dados de investigação. Esses conceitos e ações variam necessariamente consoante a disciplina e o tipo de dados produzidos, entre outros fatores".

No artigo *Digital Curation: A Life-Cycle Approach to Managing and Preserving Usable Digital Information*, Pennock (2007) justifica a abordagem do ciclo de vida através da fragilidade dos materiais digitais e dos avanços tecnológicos ao longo do seu ciclo de vida, a influência na capacidade de gerir e preservar os materiais digitais da falta de atividade nas várias etapas do ciclo de vida e a necessidade de assegurar a autenticidade e integridade dos materiais para a sua possível reutilização.

Por conseguinte, é importante que os investigadores e as instituições envolvidas em atividades de investigação disponham de competências que permitam que estes consigam gerir os conjuntos de dados (*data sets*) resultantes da investigação. Estas competências são cada vez mais importantes na medida em que permitem que os dados sejam preservados e assim sejam acessíveis e reutilizáveis por outras pessoas a longo prazo. Um dos instrumentos mais importantes é o plano de gestão de dados, que envolve todos os aspetos do ciclo de vida dos dados, tema esse que será discutido mais pormenorizadamente de seguida.

O Joint Information Systems Committee (JISC)[3] propôs um modelo do ciclo da investigação (ver Figura 2) que defende que o ciclo de vida dos dados tem origem na etapa de *processo de investigação*, etapa esta onde ocorre a criação de dados simulacionais, experimentais ou observacionais, e por isso, esta etapa não pode ser considerada à parte do ciclo de vida de investigação, já que os dados são a base para qualquer resultado científico. Segue-se a etapa de gestão de dados, que inclui a sua organização, descrição e armazenamento e assegura a qualidade dos dados. A *análise de dados* é a atividade que tipicamente acompanha a obtenção de resultados. Por fim, a *partilha de dados* envolve a curadoria de dados e tem como objetivo a preservação dos dados para que estes estejam acessíveis ao público em geral e sejam reutilizados noutros projetos.

Em suma, verifica-se através do modelo do JISC que o ciclo de vida dos dados é intrínseco aos projetos de investigação e consequentemente questões relativas à gestão, análise e partilha de dados são componentes importantes deste ciclo.

[3] Modelo proposto pelo Joint Information Systems Committee http://webarchive.nationalarchives.gov.uk/20140702225057/http://www.jisc.ac.uk/whatwedo/campaigns/res3/jischelp.aspx#ideas

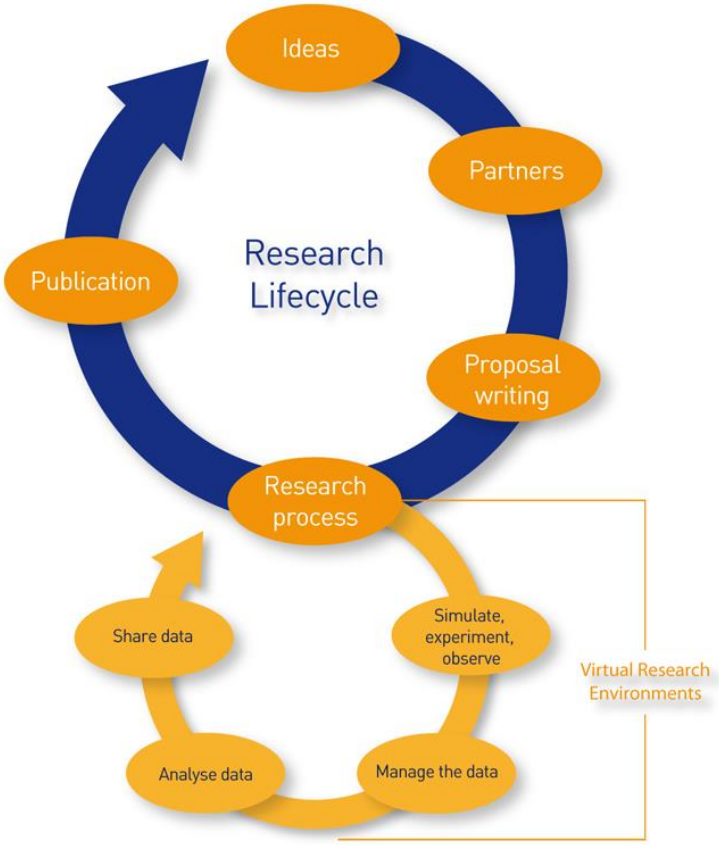


Figura 2 Ciclo de vida da investigação.

Fonte: Joint Information Systems Committee 2014

Pennock (2007) com base na justificação da necessidade de abordar o ciclo de vida, apresenta um modelo identificado na Figura 3, que tem como principal característica em comparação com outros modelos, como o apresentado anteriormente, a identificação de uma fase denominada "uso ativo" em que os dados podem estar numa base de dados onde podem ser acedidos e reutilizados para criar novos dados.

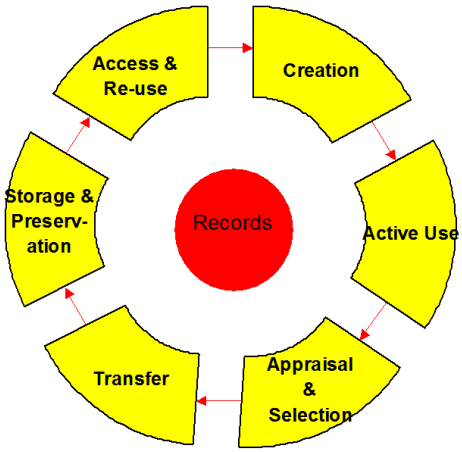


Figura 3 Ciclo de vida de dados.

Fonte: Pennock 2007

2.1.2 Planos de Gestão de Dados

Um Plano de Gestão de Dados (PGD) segundo a Harvard Library[4] (2018) "é um documento formal que descreve o que o investigador fará com os dados recolhidos durante e após um projeto de investigação". O objetivo deste plano é abordar os vários aspetos da gestão de dados, da criação de metadados e da preservação dos dados, tendo como principal foco a garantia da correta gestão dos dados.

O Digital Curation Centre (DCC)[5] também apresenta apresenta os PGD, afirmando "os planos geralmente indicam que dados serão criados e como serão criados e descrevem os planos de partilha e preservação dos dados, observando o que é apropriado dada a natureza dos dados e as restrições que possam ser necessárias"(Digital Curation Centre 2018a).

Segundo Simms and Jones (2017) o PGD "é o principal meio para cumprir as políticas de acesso aberto aos dados de investigação, e sendo a investigação uma iniciativa global, a consciência dos PGD espalhou-se para toda a comunidade de investigação".

O PGD pode ser uma solução para muitos problemas de gestão de dados e para Furtado, Príncipe, and Carvalho (2017) a elaboração de um plano pode contribuir para um trabalho mais eficaz, no sentido em que este pode simplificar o processo de investigação. O desenvolvimento de um plano, permite que os problemas sejam resolvidos com facilidade ou evitados ao planear antecipadamente a gestão dos dados.

A Harvard Library (2017) defende que muitos dos investigadores recolhem dados baseados em planos que estruturam mentalmente, no entanto esses planos ou não passam para o papel ou são muitas vezes documentados inadequadamente.

O RCAAP também se pronuncia acerca deste assunto e justifica a necessidade de desenvolver um PGD, pois este defende que o plano permite a compreensão dos dados; evita a duplicação dos dados; fortalece a continuidade do projeto de investigação; contribui para o aumento da visibilidade da investigação e para a reutilização e citação dos dados (Furtado, Príncipe, and Carvalho 2017).

Perante isto, a Harvard Library (2017) sugere que um PGD pode conter:

- Os tipos de dados que serão produzidos no processo de investigação - Qual a fonte dos dados? Em que formatos estão os dados?
- Detalhes relacionados com os metadados - Como serão documentados e descritos os dados?

---

[4] Harvard Library: https://guides.library.harvard.edu/dmp, acedido em 20 de dezembro de 2017

[5] Digital Curation Centre: http://www.dcc.ac.uk/resources/data-management-plans, acedido em 20 de dezembro de 2017

- Armazenamento, backups e políticas de segurança – Como e onde serão armazenados os dados e como serão protegidos?
- Políticas de acesso, privacidade e partilha – Questões de privacidade e confidencialidade que devem ser abordadas? Como estão disponíveis os dados?
- Políticas de reutilização – Como outros investigadores podem utilizar os dados?
- Arquivar os dados – De que forma os dados serão preservados? E quais as medidas a tomar para que estejam acessíveis a longo prazo?

Com a consciencialização da necessidade de desenvolver PGD, as agências financiadoras de projetos de investigação cada vez mais os exigem. Por conseguinte, foram desenvolvidas plataformas que auxiliam na criação de PGD. O DMPonline e DMPTool foram desenvolvidos no Reino Unido e nos EUA, respetivamente, no ano de 2011 com o intuito de apoiar as políticas de dados abertos.

- O **DMPonline**[6] é uma ferramenta online, delineada para ajudar investigadores e outros financiadores, a criar, armazenar e atualizar várias versões de planos de gestão de dados personalizados. A ferramenta disponibiliza informações que servem de orientação ao nível da gestão de dados e permite ainda exportar planos em vários formatos.
- O **DMPTool**[7] é igualmente uma ferramenta online para criar e editar um PGD. Esta ferramenta foi desenvolvida em colaboração com várias universidades e é focada especialmente em agências de financiamento dos EUA, permitindo criar PGD para agências específicas, atender aos requisitos e obter orientações para os PGDs.

De acordo com Simms and Jones (2017), o *Digital Curation Centre* (DCC) juntamente com a *University of California Curation Center*[8] estão a desenvolver a ferramenta de planeamento de gestão de dados **DMP Roadmap**, que tem como objetivo aproveitar o trabalho de projetos DMP para beneficiar toda a comunidade. Esta ferramenta para além das características do DMPOnline permitirá o trabalho colaborativo.

Para além destas ferramentas, existem também modelos desenvolvidos para criação de planos de gestão de dados, como é o caso da *University Oxford*[9] que disponibiliza um plano

---

[6] DMPonline: https://dmponline.dcc.ac.uk/, acedido em 21 de dezembro de 2017
[7] DMPTool: https://dmptool.org/, acedido em 21 de dezembro de 2017
[8] University of California Curation Center: https://www.cdlib.org/uc3/, acedido em 21 de dezembro de 2017
[9] University Oxford: http://researchdata.ox.ac.uk/portfolio/dmp-templates/, acedido em 21 de dezembro de 2017

de gestão de dados simples, adequado para projetos de investigação de pós-graduação e da *University of Virginia Library*[10] que disponibiliza exemplos de PGDs de várias disciplinas científicas.

2.2 Curadoria de Dados

Tal como acontece com a gestão de dados, a curadoria digital também tem sido alvo de atenção e de grande visibilidade e o reconhecimento da capacidade da partilha de dados científicos é cada vez maior. Neste sentido, é reconhecido que o papel do curador e do investigador são essenciais para garantir a consistência, a validade e anotações dos conjuntos de dados.

A curadoria digital garante que os dados sejam preservados, estejam disponíveis e acessíveis para que possam ser reutilizados por todas as pessoas que tenham interesse na sua utilização.

Segundo Rodrigues et al. (2010) a nível da exploração de dados existem, várias técnicas e ferramentas que facilitam a análise dos dados. No entanto, quando se refere à curadoria e à partilha de dados, entre várias áreas e disciplinas científicas, o mesmo não acontece. A explicação está na diversidade de culturas das comunidades científicas e nas infraestruturas.

"Algumas disciplinas, como a genética, as ciências climáticas e a astronomia, possuem infra-estruturas e normas bem implantadas que facilitam a pesquisa, acesso e reutilização dos dados científicos, enquanto noutras falta ainda quase tudo para que a partilha de dados se possa realizar de forma generalizada" (Rodrigues et al. 2010).

Para Pennock (2007), "a curadoria digital, amplamente interpretada, consiste em manter e agregar valor a um corpo confiável de informações digitais para utilização atual e futura: noutras palavras, é a gestão ativa e avaliação da informação digital". Segundo o (Digital Curation Centre 2018c) este processo implica manter, preservar e adicionar valor aos dados de investigação ao longo do ciclo de vida.

Para que a partilha e a reutilização de dados se possa realizar é importante que existam profissionais que estejam focados na curadoria. Os investigadores, mais focados na investigação, não são as pessoas ideais para assegurar a preservação e o acesso contínuo dos dados. É neste sentido, que o papel do curador surge, ligado a profissionais na área da informação ou que tenham alguma experiência e que integrem os grupos de investigadores.

---

[10] University of Virginia Library: http://data.library.virginia.edu/data-management-plan-templates/, acedido em 21 de dezembro de 2017

O Digital Curation Centre (DCC) um dos mais importantes centros de curadoria digital no artigo "*Curating E-Science Data*" apresenta os benefícios da curadoria digital, para os intervenientes no processo de investigação (criadores de dados, investigadores, financiadores e investigadores que pretendam reutilizar os dados), a curto e longo prazo.

Segundo o DCC, a curadoria digital a curto prazo permite a aplicação de padrões de captura de dados e melhoria da qualidade dos dados; permite que grandes quantidades de dados sejam analisadas e que investigação relacionada esteja ligada e que os *data sets* possam ser planeados de forma inovadora, permitindo o acesso a dados confiáveis.

A longo prazo também são identificados alguns benefícios pelo DCC, nomeadamente, a maximização do potencial dos dados recolhidos e analisados; disponibilização de estruturas que permitem migrar os dados de forma controlada; permite a verificação da proveniência dos dados; garante a preservação dos dados que não é possível reproduzir e ainda permite que os investigadores que reutilizam os dados o façam de forma confiável.

Para Pennock (2007) a curadoria digital também enfrenta desafios, nomeadamente o tipo de informação a tratar, os recursos que os investigadores têm disponíveis, a fase em que uma instituição assume a responsabilidade pelos recursos, e as políticas da instituição. No entanto, o autor afirma que apesar disto, atividades de diferentes domínios podem ser úteis para outros, quer a nível de técnicas, quer a nível das lições aprendidas e dos resultados obtidos que podem evitar a duplicação de trabalho.

Para o RCAAP,

> "A curadoria dos dados não se esgota obviamente na criação de metadados e compreende o conjunto das acções que garantem a autenticidade, integridade e acessibilidade dos dados científicos. Em especial, a curadoria dos dados envolve todas as actividades de preservação necessárias para garantir a possibilidade de voltarem a ser usados no futuro" (Rodrigues et al. 2010).

Com o objetivo de criar uma ferramenta que auxiliasse os curadores de dados a entender todas as atividades envolvidas na curadoria, o *Digital Curation Centre* desenvolveu um modelo[11], apresentado na Figura 4, que representa o ciclo de vida da curadoria digital,

[11] Modelo do Ciclo de vida da curadoria: http://www.dcc.ac.uk/resources/curation-lifecycle-model, acedido em 21 de dezembro de 2017

onde expõe uma visão geral das fases necessárias para a curadoria desde a recolha dos dados até à preservação.

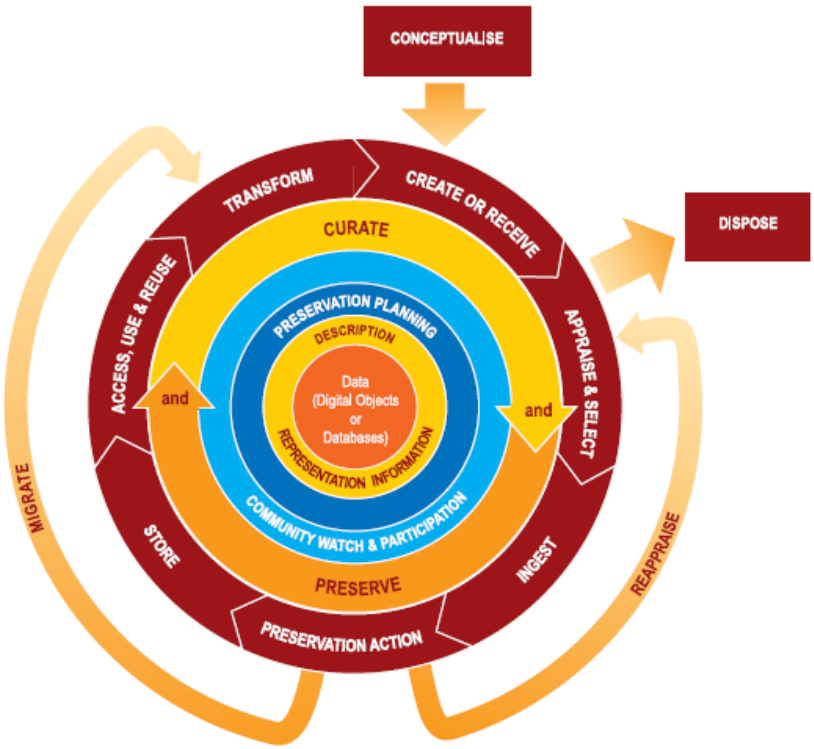


Figura 4 Ciclo de Vida da Curadoria Digital.

Fonte: Digital Curation Centre 2018b

O Digital Curation Centre (2018c) apresenta detalhadamente as várias etapas do ciclo de vida de curadoria digital, que podem ser sintetizadas em:

- **Planeamento** - planeamento da criação dos objetos digitais, nomeadamente os métodos de recolha de dados;
- **Criação** - produção de objetos digitais e atribuição de metadados;
- **Acesso, utilização e reutilização** - garantir o acesso aos objetos digitais para que estes possam ser reutilizados noutros projetos;
- **Avaliação e seleção** - avaliar os objetos digitais e selecionar aqueles que se destinam a passar pelo processo de curadoria e preservação;
- **Eliminação** - eliminar objetos que não foram selecionados para preservação a longo prazo, seguindo políticas e requisitos legais;
- **Ingestão** - transferir os objetos digitais para um repositório, segundo políticas e requisitos legais;
- **Preservação** - garantir a preservação a longo prazo dos objetos digitais;
- **Reavaliação** - reavaliar os objetos digitais que falharam nos processos de validação;

- **Transformação** - criar novos objetos digitais a partir dos originais (por exemplo: migração).

As etapas referidas anteriormente são fundamentais para que a curadoria dos dados seja efetuada de forma eficiente. Isto leva a que, os dados de investigação possam ser armazenados, partilhados e reutilizados de forma a gerar novo conhecimento científico, a diminuir o esforço de duplicação dos dados e aumentar o valor dos dados. Assim, o ciclo de vida da curadoria digital apresenta-se como um método fundamental para os curadores de dados, que devem seguir todas as etapas de forma a garantir a gestão ativa dos dados e consequentemente a reduzir as ameaças de obsolescência tecnológica.

Por conseguinte, estratégias de curadoria de dados, preservação e armazenamento são partes constituintes da gestão de dados de investigação e são fundamentais para que os dados sejam interpretáveis pelas comunidades de investigadores. A curadoria de dados está bem marcada nas comunidades científicas através dos vários repositórios temáticos e dos repositórios generalistas, que tentam aglomerar vários domínios e dos centros de dados que tentam orientar os profissionais e disponibilizar ferramentas que facilitem a realização desta atividade.

Rodrigues et al. (2010) alerta para o problema das "pequenas ciências" "(…) com os conjuntos de dados produzidos por investigadores individualmente ou por pequenos grupos de investigação que não possuem nem os recursos nem as infra-estruturas para cuidar dos conjuntos de dados" e apresenta os repositórios institucionais como resposta à necessidade da curadoria.

2.3 Metadados para Dados de Investigação: modelos de metadados

Tal como referido anteriormente, para o Digital Curation Centre (2018c) o processo de curadoria envolve " agregação de valor" aos dados de investigação ao longo do seu ciclo de vida. É neste sentido que surgem os metadados, como forma de adicionar valor, informação importante sobre os dados.

Segundo Higgins (2018) "metadados são informação descritiva ou contextual que se refere e está associada a outro objeto ou recurso". A autora realça ainda que "sem eles um recurso digital pode ser irrecuperável, não identificável ou inutilizável".

Pode-se então dizer que os metadados são dados sobre os dados, que explicam em que contexto foram criados, onde e quando e por isso os dados devem ser descritos o mais

detalhadamente possível garantindo que são utilizados metadados de qualidade, ou seja, que cumpram a sua intenção, garantir que os dados sejam compreendidos.

Desta forma, é importante descrever os dados para que este possam ser facilmente encontrados, para que sejam corretamente preservados a longo prazo e para que sejam partilhados e reutilizados por outros, pois se os dados estiverem descritos corretamente, vão ser entendidos por outros investigadores, por outras palavras a descrição de metadados garante a preservação, recuperação e reutilização dos dados.

Para Ribeiro et al. (2016) os metadados também são valiosos antes dos dados serem depositados, pois os investigadores partilham dados entre projetos e o facto de estes terem descrições agregadas, são mais úteis, pois permitem que sejam interpretados mais facilmente e por isso a cooperação entre investigadores seja simplificada.

Relativamente aos dados de investigação, que serão o foco deste trabalho, estes são produzidos durante experiências que posteriormente serão interpretadas, neste sentido, é ainda mais importante que os dados contenham o máximo de informação possível para ajudar a sua compreensão.

Segundo Higgins (2018) os metadados podem ser categorizados de acordo com as funções, assim sendo identificam-se quatro categorias de metadados:

- **Metadados descritivos:** permitem a identificação, localização e recuperação dos dados pelos utilizadores e podem incluir utilização de vocabulários controlados, indexação, entre outros.
- **Metadados técnicos:** descrevem os processos técnicos utilizados para produzir os dados.
- **Metadados administrativos:** descrevem aspetos relacionados com os direitos de propriedade intelectual e aquisição e ainda com a criação, alteração e controlo de versões dos metadados**.**
- **Metadados de preservação:** documentam ações realizadas para preservar os dados, por exemplo: migrações.

Ao longo dos tempos foram apresentados diferentes esquemas de metadados com o objetivo de normalizar a descrição, através de estruturas formais.

O *DublinCore* (DC)[12] é um esquema de metadados mais geral utilizado para recursos na web e outros. No entanto, existem outros mais específicos, como é o caso do domínio

[12] DublinCore: http://dublincore.org/, acedido em 21 de dezembro de 2017

ambiental, de esquemas relacionados com Internet das coisas (Internet of Things—IoT), sensores meteorológicos, que podem ser importantes para os *data sets* analisados neste projeto, sendo alguns exemplos desses esquemas apresentados de seguida.

Os modelos de metadados apresentados podem ser utilizados numa fase posterior do trabalho como solução para a aplicação de metadados adequados aos dados analisados, em conjunto com outros que poderão vir a ser desenvolvidos.

2.3.1 Dublin Core

O Dublin Core (DC)[13] é um esquema de metadados desenvolvido para descrever recursos digitais na web, não está associado a nenhum domínio específico e é um dos esquemas mais conhecidos.

O DC pode ser facilmente compreendido e implementado por qualquer pessoa, mesmo sem que esta tenha algum tipo de conhecimento ou experiência na área de indexação. Este modelo de metadados é composto por quinze elementos genéricos que podem ser utilizados para descrever quase todo o tipo de recursos disponíveis na web. Os elementos que compõem o vocabulário Dublin Core e que fazem parte de um conjunto maior de vocabulários da *Dublin Core Metadata Initiative* (DCMI) são "title"; "creator"; "subject"; "description"; "publisher"; "contributor"; "date"; "type"; "format"; "identifier"; "source"; "language"; "relation"; "coverage" e "rights". Para além destes elementos, o DC permite que sejam adicionados mais elementos de acordo com as necessidades das comunidades.

2.3.2 Data Documentation Initiative (DDI)

O *Data Documentation Initiative* (DDI)[14] é um padrão internacional utilizado para descrever dados das ciências sociais, do tipo observacional, comportamental, económico e da saúde recolhidos em pesquisas.

Atualmente existem duas versões disponíveis, a *DDI Codebook* publicada em janeiro de 2012 e modificada em janeiro de 2014, que se trata de uma versão mais simples. Esta versão foi designada para documentar dados de investigação simples e preparada para facilitar a migração de documentos para o *DDI Lifecycle*, outra das versões disponíveis, esta mais desenvolvida que auxilia a descrição dos dados ao longo de todo o ciclo de vida.

---

[13] Metadados Dublin Core: http://dublincore.org/documents/dcmi-terms/, acedido em 21 de dezembro de 2017

[14] Data Documentation Initiative (DDI): https://www.ddialliance.org/, acedido em 21 de dezembro de 2017

2.3.3 Friend of a Friend (FoaF)

A *Friend of a Friend* (FoaF)[15] é uma ontologia interpretável por máquinas que tem como objetivo ligar pessoas e informações através da Web, permitindo a descrição de redes sociais sem a necessidade de uma base de dados centralizada.

A FoaF abrange três tipos de redes: redes sociais, redes representacionais e redes de informação. São exemplos, alguns dos elementos que compõem a ontologia FoaF: "Agente" corresponde a uma pessoa, grupo ou programa, "Project" corresponde a um projeto formal ou informal, coletivo ou individual.

2.3.4 Ontology for Meteorological Sensors (aws)

A *Ontology for Meteorological Sensors* (aws),[16] desenvolvida pelo grupo *W3C Semantic Sensor Networks,* é uma ontologia que pertence ao domínio da Meteorologia da Agricultura e que apesar de não ser uma ontologia generalista, pode ser vista como uma solução a adotar juntamente com outros modelos de metadados, para a realização deste trabalho. O vocabulário aws é composto por vários elementos, sendo alguns exemplos disso, o "SensingDevice" que descreve um dispositivo de deteção, o "CapacitanceSensor" corresponde à capacidade dos sensores em relação à humidade quando sofrem mudanças causadas pela interação com vapor de água e o "CapacitiveAluminumOxide" descreve também a capacidade dos sensores em relação à sua oxidação.

2.3.5 M3-lite

O M3-lite [17]é uma versão da ontologia M3, desenvolvida para o projeto EU H2020 FIESTA-IoT[18], que permite anotar semanalmente os dados IoT produzidos por diferentes dispositivos. Esta taxonomia abrange vários domínios, entre os quais, a classificação de dispositivos, fenómenos e unidades de medida, fenómenos físicos e ambientais, diferentes tipos de sensores, entre outros. Alguns dos elementos que compõem o vocabulário do M3-lite são o "calibration measurement typec" utilizado quando os dados são recolhidos enquanto os sensores estão a ser calibrados; o "coordinates" que descreve a localização ao nível da latitude, longitude e altitude; "day" descreve o dia em que foram recolhidos os

---

[15] Friend of a Friend (FoaF): http://xmlns.com/foaf/spec/, acedido em 21 de dezembro de 2017

[16] Ontology for Meteorological Sensors: https://www.w3.org/2005/Incubator/ssn/ssnx/meteo/aws, acedido em 21 de dezembro de 2017

[17] M3-lite: http://ontology.fiesta-iot.eu/ontologyDocs/fiesta-iot/doc, acedido em 21 de dezembro de 2017

[18] FIESTA-IoT: http://fiesta-iot.eu/, acedido em 21 de dezembro de 2017

dados; "degree celsius" descreve a temperatura em graus celsius e o "device" que pode ser utilizado para descrever um dispositivo tecnológico utilizado na recolha dos dados.

2.3.6 Climate and Forecast (CF)

O *Climate and Forecast* (CF) [19] é um modelo de metadados patrocinado pela Convenção de Metadados de Clima e Previsão da NetCDF[20]. Inicialmente este esquema foi desenvolvido para dados escritos em formato netCDF, ou seja, dados que estão independentes do sistema que sustentam a sua criação, acesso e partilha, mais direcionados para dados de previsão do tempo. Contudo, este modelo de metadados pode ser aplicado a conjuntos de dados observacionais e pode ser utilizado noutros formatos para além do NetCDF. Este modelo de metadados pode ser utilizado para localizar os dados no espaço-tempo e para descrição física, sendo alguns exemplos dos elementos que o compõem, a "altitude"; "plataform_id" ou "plataform_name" que ajudam a identificar a plataforma a partir da qual uma observação foi feita, entre outros.

2.3.7 Semantic Sensor Network Ontology (SSN)

O *Semantic Sensor Network Ontology* (SSN) [21]é uma ontologia desenvolvida pelo W3C em colaboração com o Open Geospatial Consortium (OGC)[22] utilizada para definir um esquema geral para a descrição de sensores, da sua precisão e as capacidades dos mesmos.

A participação do OGC neste projeto, é um aspeto distintivo da ontologia SSN, uma vez que o projeto trabalha com o objetivo de tornar o SSN compatível com os padrões OGC existentes, nomeadamente, *OGC Observation and Measurements standard,* um vocabulário XML que descreve o modelo concetual de Observation and Measurements (O&M) na publicação de observações ambientais e o *OGC Sensor Model Language (SensorML)* que é também um vocabulário XML utilizado para descrever sensores e processos de medição.

Esta ontologia pode ainda ser utilizada na descrição do tempo de implementação e duração do sensor. Alguns dos elementos que compõem o vocabulário SSN e que poderão

---

[19] Climate and Forecast: http://www.dcc.ac.uk/resources/metadata-standards/cf-climate-and-forecast-metadata-conventions

[20] NetCDF: https://www.unidata.ucar.edu/software/netcdf/

[21] Semantic Sensor Network Ontology: https://www.w3.org/2005/Incubator/ssn/ssnx/ssn#hasRegion

[22] Open Geospatial Consortium: http://www.opengeospatial.org/

ser utilizados na descrição dos *data set* analisados neste projeto são "region"; "Method"; "Process"; "InformationObject", entre outros.

2.4 Formatos de representação de dados

Relacionada com a questão da criação de metadados está também a questão da representação dos dados na Web Semântica.

É neste sentido que foram criadas linguagens que permitem a partilha de informação estruturada numa aplicação. As linguagens são tecnologias de representação, utilizadas de forma geral para se conseguir perceber a natureza dos metadados, e o significado de cada descritor utilizado.

De entre a coleção de tecnologias de dados existentes, que fornecem um ambiente onde é possível consultar os dados e extrair inferências usando vocabulários, são apresentadas as três linguagens de representação, *Extensible Markup Language (XML)*, *Resource Description Framework (RDF)* e *Web Ontology Language (OWL)*.

2.4.1 XML (Extensible Markup Language)

Extensible Markup Language (XML)[23] foi desenvolvido pelo *Standard Generalized Markup Language Editorial Review Board* e consiste numa linguagem simples, que permite descrever vários tipos de dados organizados de forma hierárquica. Esta linguagem é aplicada para construir documentos semiestruturados na Web e pode ser utilizada para expor um modelo de metadados.

O XML fornece um framework regular e um conjunto de ferramentas de análise sintática e troca de dados/metadados entre aplicações, ou seja, permite que os dados que estão na Web sejam acedidos por máquinas. Os documentos XML são compostos por elementos de armazenamento denominados de "entidades", que incluem dados analisados ou não analisados. Os dados analisados são constituídos por caracteres , sendo que alguns destes formam dados de caracteres outros formam marcações.

2.4.2 RDF/ XML

O Resource Description Framework (RDF)[24] é um conjunto de especificações reservadas à definição de relações entre objetos, utilizada para descrever e relacionar

[23] Extensible Markup Language: https://www.w3.org/TR/REC-xml/#sec-origin-goals
[24] Resource Description Framework: https://www.w3.org/TR/rdf-concepts/

recursos na Web. Esta *framework* segue uma estrutura de declarações sobre recursos, constituído por um sujeito, predicado (que também pode ser chamado de propriedades) e um objeto. O assunto e o objeto representam os dois recursos que estão a ser relacionados e o predicado descreve a relação entre os dois anteriores.

O RDF dispõe de um domínio independente, ou seja, é responsabilidade do utilizador definir o vocabulário que pretende utilizar, esta tarefa pode ser feita através de uma linguagem denominada RDFSchema.

Esta *framework* tem uma vantagem, que a destaca de outras, que é a flexibilidade, ou seja, independentemente do número de objetos, as relações podem ser descritas facilmente.

2.4.3 OWL (Ontology Web Language)

Ontology Web Language (OWL)[25] foi criada com o objetivo de ser utilizada para processar informações contidas nos documentos utilizando aplicações e pode ser utilizada para representar o significado de termos e de relações entre esses termos em vocabulários. Esta linguagem acrescenta significado a um modelo de metadados, permite definir ontologias na web utilizando triplos.

A OWL suporta a criação de vocabulários adicionais para descrever propriedades e classes, relações entre classes, cardinalidade, igualdade, características de propriedades entre outros recursos, que promovem as características necessárias para representações de ontologias mais complexas. Esta linguagem OWL utiliza a mesma sintaxe da linguagem RDF.

2.5 Preservação de dados de investigação

A curadoria de dados implica todas as atividades de preservação necessárias para garantir a possibilidade de os dados poderem ser utilizados no futuro por outros investigadores ou por todos aqueles que tenham interesse neles.

Segundo Furtado, Príncipe, and Carvalho (2017), entende-se por preservação digital o "conjunto de medidas e infraestruturas necessárias para garantir a integridade de dados digitais a longo-termo, por vários anos e idealmente por tempo a termo indeterminado (*ad eternum*).".

Em geral, a preservação digital tem como objetivo garantir o acesso contínuo e ativo, garantindo a durabilidade do objeto mesmo perante os desafios da obsolescência

[25] Ontology Web Language: https://www.w3.org/TR/2004/REC-owl-features-20040210/#s1

tecnológica. No entanto Bicarregui et al. (2013) colocam uma questão importante, quais os objetivos da preservação? Os autores dizem que não se deve pensar em preservar tudo, pois essa tarefa trás consigo custos, que nem sempre podem ser suportados, a resposta está nas necessidades de cada disciplina.

Quando se pensa em dados de investigação e na sua preservação, a problemática é a mesma, assim como as soluções. Rodrigues et al. (2010) defendem que a necessidade de preservação dos dados é comum a todas as áreas, no entanto, o tipo de dados a preservar e o período de preservação pode variar, assim como as estratégias a adotar. Segundo os autores as estratégias de preservação escolhidas variam com muitos fatores, nomeadamente, as características dos dados e a existência de formatos de dados e esquemas de metadados adequados para preservar os dados, a existência de infraestruturas de armazenamento e as capacidades técnicas.

A preservação é uma medida importante na gestão de dados, pelo que a descrição adequada dos dados é uma tarefa essencial. Para isso é importante a utilização de metadados ricos que descrevam por exemplo, a proveniência dos dados. Apesar do tempo que tarefas como a descrição ocupam, estas são justificadas pois trazem benefícios para a preservação de dados quando realizadas com ferramentas adaptadas. Esta tarefa juntamente com a seleção deve fazer parte do processo de gestão de dados desde o início.

Neste sentido, é importante que os investigadores tomem consciência desde logo de algumas tarefas que devem ter em conta, nomeadamente, a definição do ciclo de vida dos dados recolhidos na investigação, apresentando os prazos relacionados com armazenamento em repositórios, a preservação e a partilha de dados; devem ainda identificar os repositórios onde irão armazenar os dados. É de referir que os primeiros passos quanto à preservação de dados de investigação deverão ser dados ainda numa fase de planeamento, através da elaboração de um Plano de Gestão de Dados.

Verifica-se que há uma maior consciência no que respeita à necessidade de armazenamento e de preservação de dados em repositórios criados para corresponder a esse objetivo. Independentemente do repositório, a preservação de dados a longo prazo deve ser convenientemente refletida.

Os repositórios surgem como infraestruturas que regulam o processo de ingestão, arquivo, gestão, preservação e acesso aos dados. Deste modo, os responsáveis pelos repositórios têm uma função importante no que respeita à preservação, a qual deve garantir serviços de qualidade. Assim sendo, é fundamental que os responsáveis dos repositórios assegurem a robustez e a fiabilidade dos mesmos, a partilha de protocolos a serem utilizados pelos investigadores para regular o arquivo dos dados no repositório, entre outros.

O armazenamento dos dados num repositório digital é importante para garantir a integridade dos dados a longo prazo, utilizando para isso várias medidas de preservação, como é o caso dos formatos. É essencial garantir que os dados sejam armazenados num formato confiável. A preservação nos repositórios é um processo contínuo em que os dados necessitam de ser geridos e por isso é importante que os metadados sejam armazenados juntamente com os dados que eles descrevem, para evitar perdas dos metadados e que os dados sejam armazenados de forma segura, para evitar que sejam modificados (os metadados permitem ver se os dados foram alterados em algum momento).

As medidas de preservação de dados são modificadas e melhoradas ao longo do tempo. Por exemplo, a da utilização de formatos de arquivo uniformes, ou seja, colocar o conjunto de dados num só formato; atualizações frequentes de dados através de cópias de dados; migração de dados de formatos mais antigos para formatos mais recentes e emulação de *software* descontinuados para aceder a formatos de dados antigos.

3 Abordagem metodológica

As metodologias utilizadas para o projeto de dissertação são fundamentais para o sucesso da sua realização e obtenção de resultados, já que estas são metodologias importantes para o trabalho de investigação e podem influenciar a tomada de decisão do investigador.

A metodologia utilizada na dissertação como se pode ver na Figura **5** inicia-se com o processo de contextualização do projeto através da análise de artigos e documentos já publicados acerca do tema retratado, o que permite fundamentar o trabalho desenvolvido e contribui para uma maior veracidade do trabalho. Já com algum conhecimento do projeto e foi feito o *download* dos dados através da plataforma UrbanSense.

Uma vez que o projeto não possui nenhum Plano de Gestão de Dados criá-lo foi uma tarefa inicial importante, uma vez que o PGD serviu de linha orientadora da gestão dos dados.

Posteriormente foi necessário organizar os dados em pastas de acordo com as estações e os sensores a que pertencem e ainda foi criado um código de referência de forma a uniformizar os conjuntos de dados. Uma vez que os dados analisar pertencem a um domínio não conhecido pelo autor, é importante a realização de entrevistas semiestruturadas aos investigadores que produziram os dados e que estão a analisar os dados da plataforma UrbanSense. Com estas entrevistas pretende-se entender melhor os dados que serão analisados e também como é realizado todo o processo de recolha, uma vez que esta implica a utilização de sensores especializados. A realização de entrevistas permite ainda compreender os utilizadores que podem demonstrar interesse em aceder e utilizar os dados.

A adoção do tipo de entrevista, neste caso, semiestruturada recaiu sobre a sua flexibilidade que não implica a adoção de protocolos fixos. Como é característico da entrevista semiestruturada, é previamente estabelecido um conjunto de questões. No entanto, ao longo da entrevista são incluídas outras questões que podem surgir.

Depois de compreender melhor o processo de recolha dos dados e o projeto mais detalhadamente, analisaram-se elementos de vários modelos de metadados e de entre todos os metadados analisados foram selecionados alguns. Depois de selecionados os metadados, foi necessário formalizar os modelos de metadados M3-lite e SSN (anteriormente selecionados)

numa ontologia e iniciou-se o processo de descrição dos *data sets*, utilizando para isso a plataforma de gestão de dados — Dendro.

A última etapa no *workflow* de gestão de dados corresponde ao depósito dos dados no repositório do INESC TEC.

Esta dissertação é ainda composta por duas fases de avaliação, a primeira das quais será feita antes do preenchimento dos metadados. Após a seleção dos descritores utilizados na descrição dos *data sets*, estes serão enviados ao investigador principal do projeto UrbanSense, para que o investigador faça uma avaliação dos descritores selecionados.

Numa segunda fase, depois da descrição e publicação dos *data sets*, será ainda feita uma nova avaliação junto de dois investigadores, um investigador do domínio e outro investigador que não pertence ao domínio mas que está familiarizado com a curadoria de dados. Para proceder a essa avaliação os investigadores acederam ao repositório do INESC TEC, onde estavam publicados os dados, foi-lhes explicado o contexto desta dissertação e dessa avaliação. No entanto, nada foi explicado acerca do projeto UrbanSense de forma a não enviesar os comentários dos investigadores.

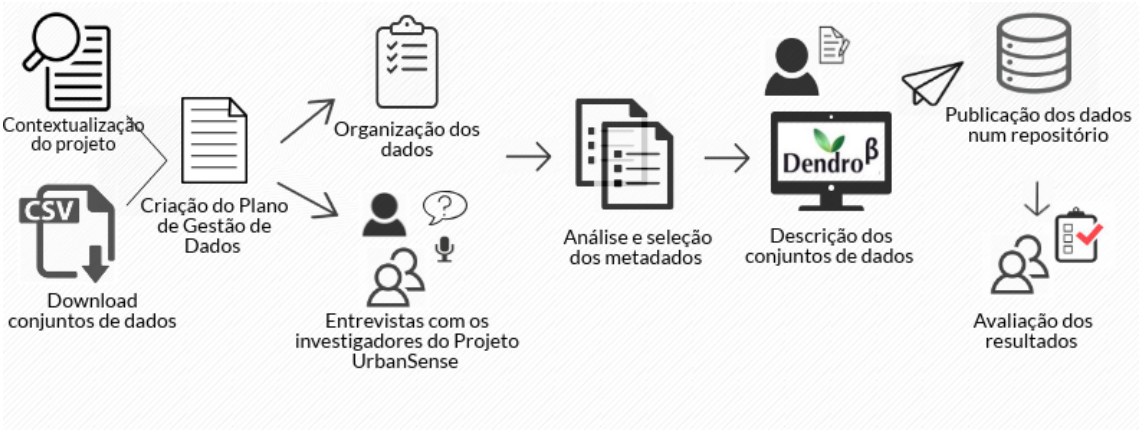


Figura 5 Workflow de Gestão dos Dados UrbanSense.

4 Plataforma para Smart Cities

O paradigma da Internet of Things (IoT) tem apoiado o setor das Smart Cities, uma vez que estas se tornaram um meio para o desenvolvimento de aplicações que visam melhorar a qualidade de vida dos cidadãos e produzir informação sobre os processos urbanos, que podem desencadear o desenvolvimento de novos serviços para os cidadãos com base no conhecimento adquirido a partir das Smart Cities. Áreas como a qualidade do ambiente, a mobilidade, a segurança e a saúde são áreas em que a IoT tem desempenhado um papel mais relevante.

Tendo em conta que as cidades são sistemas complexos que aglomeram vários processos, como serviços, recursos, sistemas sociais, entre outros, as tecnologias de sensores têm constituído cada vez mais a forma de concretizar este paradigma, tornando os sistemas de informação cada vez mais incorporados no ambiente que nos rodeia, de uma forma quase invisível. A aplicação da IoT no contexto urbano, como é o caso do projeto PortoLivingLab, é cada vez mais relevante pois surge como como resposta à adoção de soluções de Tecnologias de Informação e Comunicação (TIC) nas cidades, as chamadas Smart Cities.

Como resultado deste desenvolvimento são produzidas grandes quantidades de dados que são recolhidos através de dispositivos sensoriais e que monitorizam dados relacionados com a temperatura, luz, movimento, entre outros e que necessitam de ser armazenados, analisados e disponibilizados, para que sejam facilmente interpretáveis.

O PortoLivingLab é um exemplo de uma infraestrutura de adoção de soluções tecnológicas implementada na cidade do Porto, que aproveita a tecnologia IoT para compreender a cidade através de quatro fenómenos: clima, meio ambiente, transportes públicos e fluxos de pessoas.

Esta infraestrutura, que recolhe vários tipos de dados, é composta por três plataformas de monitorização como é possível ver na Tabela 1, que atuam em paralelo desde 2013, nomeadamente,

- **SenseMyCity**, uma ferramenta móvel que recolhe dados do sensor dos dispositivos e de acesso dos participantes através de uma aplicação móvel, garantindo a privacidade e anonimidade dos dados recolhidos;
- **UrbanSense**, um conjunto de sensores implementados em locais estratégicos que recolhem dados ambientais, como por exemplo, localizações geográficas, temperatura, qualidade do ar, precipitação, luminosidade, entre outros;

- **BusNet** uma plataforma que identifica e partilha dados GPS e metadados de ligação Wi-Fi, composta por mais de 600 veículos, equipados com unidades on-board (OBUs) que permite comunicação veículo a veículo (V2V) e veículo- infraestrutura (V2I) em que os veículos estão equipados com acesso à internet para os passageiros.

Tabela 1 Resumo das plataformas do PortoLivingLab.

Fonte: Traduzido de Santos et al., n.d

	SenseMyCity	**UrbanSense**	**BusNet**
Tipo	Sensor de multidão	Sensores estáticos	Nós veiculares
Ano de lançamento	2011	2015	2014
Nº nós	677 participantes	19 DCUs	608 OBUs/RSU
Nº entradas	370M	68M	1M/dia (=700M)
Informação	Mobilidade e perspetiva dos cidadãos	Meio Ambiente	Fluxos de passageiros e concetividade
Dados não tratados	Localização, movimento, hotspots WiFi, questionários, externos (por exemplo, OBD)	Ruído, humidade, temperatura, UV, partículas, vento, chuva, luminosidade	Localização de autocarros, ligações WiFi de passageiros

O objetivo principal desta infraestrutura centra-se na identificação da dinâmica do clima, qualidade do ar, ruído e a mobilidade através dos dados recolhidos pelos vários sensores (pessoais, estáticos e de veículo) e pelo desenvolvimento de aplicações que demonstrem a utilidade desses dados.

Os dados, após serem identificados pelos sensores das plataformas IoT urbanas, são recolhidos a partir das infraestruturas espalhadas pela cidade para depois serem armazenados no servidor backend cloud tal como se pode ver na Figura 6.

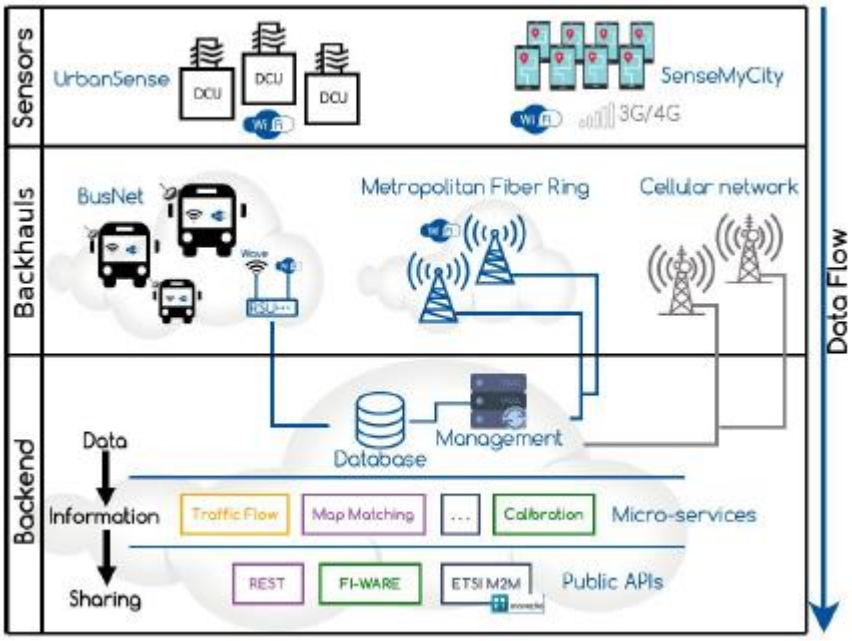


Figura 6 Arquitetura do PortoLivingLab.

Fonte: Santos et al., n.d

O fluxo de dados pode ser dividido em três etapas, a primeira etapa - "**Sensing**" - corresponde à recolha dos dados quer ambientais, no caso da plataforma UrbanSense, quer dos dados das rotas dos cidadãos que utilizam a plataforma SenseMyCity e das rotas dos veículos e dados Wi-Fi no caso da plataforma BusNet. Para recolher estes dados, a plataforma UrbanSense, que é constituída por unidades estáticas, monitoriza vários parâmetros relacionados com o clima, qualidade do ar e qualidade de vida. A plataforma SenseMyCity transforma os dispositivos móveis em sondas que detetam a localização do utilizador, os movimentos que este faz e ao mesmo tempo pode recolher o feedback dos utilizadores através de questionários, por exemplo o humor que naquele dia caracteriza o utilizador ou então o meio de transporte que o utilizador usou para chegar a determinado local. A rede BusNet, através das unidades on-board, recolhe informação sobre o trajeto do veículo e as ligações que são feitas através do AP Wi-Fi.

Na segunda etapa – "**Data Collection**" são aproveitadas as infraestruturas de comunicação complexas existentes em áreas urbanas, juntamente com uma Delay Tolerant Network (DTN), em que os sensores da infraestrutura PortoLivingLab através dos pontos de ligação sem fios alcançam os nós backhaul. No caso do projeto UrbanSense, os dados recolhidos são transferidos para o servidor backend através APs Wi-Fi, e no caso do projeto SenseMyCity os dados são transferidos através de ligação Wi-Fi ou através de comunicações telefónicas, dependendo da preferência do utilizador. Relativamente à rede BusNet a transferência dos dados é feita através das comunicações V2X e da Delay Tolerant Network.

Por fim, a última etapa do fluxo de dados - "**Storage, Processing, Sharing**" – corresponde ao armazenamento dos dados recolhidos pelos vários sensores em bases de dados. Apesar dos dados serem recolhidos a partir de várias fontes, eles têm semelhanças e passam por uma análise, filtragem e pré-processamento do serviço backend, o que permite a sua fusão. Relativamente à partilha dos dados, para entidades externas a partilha é realizada a partir de um programa middleware e ainda os dados processados são disponibilizados publicamente em tempo real através de interfaces, em que os dados de crowdsensing são partilhados anonimamente.

4.1 UrbanSense

A Plataforma UrbanSense é uma infraestrutura desenvolvida para monitorizar vários fenómenos ambientais, composta por unidades de recolha de dados (Data Collection Units -DCU), (ver Figura 7) constituídas por um conjunto de 13 sensores que estão colocados em locais estratégicos na cidade do Porto. Cada unidade é composta por vários sensores que medem tempo, ruído, qualidade do ar e luminosidade.

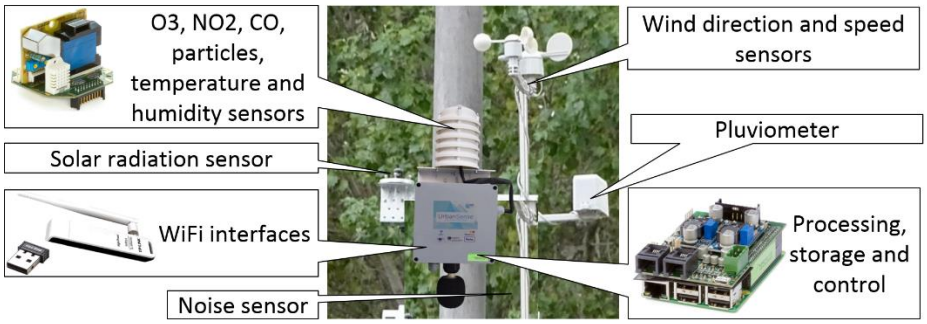


Figura 7 Aspeto externo da DCU.

Fonte: Luis et al. 2016

Segundo Luis et al. (2016), a plataforma UrbanSense é constituída por três componentes como se pode ver na Figura 8, as unidades de recolha de dados (DCU), a componente de transporte dos dados que permite a utilização de diferentes formas de comunicação para recolher os dados de todas as DCUs e por fim a componente de processamento composta por servidores *backoffice* para armazenamento de dados, dispondo de uma interface para disponibilizar os dados para clientes externos.

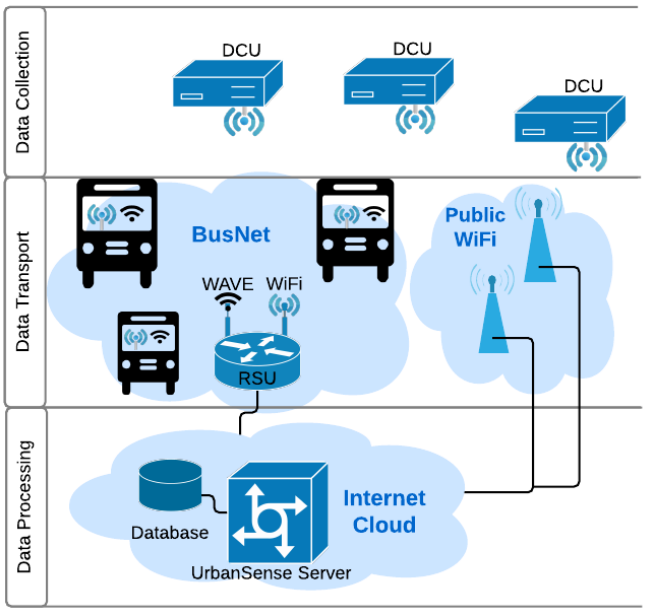


Figura 8 Arquitetura da plataforma UrbanSense.

Fonte: Luis et al. 2016

As DCUs são unidades essenciais da plataforma UrbanSense compostas por uma fonte de alimentação, uma unidade de processamento, uma placa de controlo e vários sensores (ver

Tabela **2**). Os sensores podem ser divididos em três classes:

- **Meteorológicos**, dos quais fazem parte alguns sensores como o termómetro, higrómetro (utilizados para medir a humidade na atmosfera), anemómetro (utilizado para medir a velocidade de um fluido que pode ser ar ou água) e o pluviómetro (utilizado para recolher e medir, em milímetros lineares, a quantidade de líquidos ou sólidos como por exemplo, chuva, neve ou granizo);
- **qualidade vida**, em que são usados sensores como o sonómetro (utilizado para medir os níveis de pressão sonora), medidor lux (utilizado para medir a intensidade de luz) e radiação solar;
- **qualidade do ar**, desta classe fazem parte o sensor de partículas, de CO, de NO2 e de O3 .

Tabela 2 Lista de sensores e suas características.

Fonte: Luis et al. 2016

Sensor	Digital	Bus	Heating element	Power consumption
Temperature & RH	✓	I^2C		$5mW$
Precipitation	✓	1 channel		$< 6mW$
Wind speed	✓	1 channel		$< 6mW$
Wind direction		1 channel		$< 6mW$
Luminosity	✓	I^2C		$\sim 1mW$
Sonometer	✓	UART		$< 500mW$
Solar radiation		1 channel		self-powered
Particles	✓	2 channels		$\sim 450mW$
CO		1 channel		$15-150mW$
NO_2		1 channel	✓	$< 50mW$
O_3		1 channel	✓	$< 100mW$

A implementação do projeto UrbanSense foi alvo de alguns desafios, nomeadamente no que respeita aos sensores, uma vez que o projeto está dependente de infraestruturas e equipamentos de instituições parceiras e pelo facto dos equipamentos sofrerem desgaste devido às condições climáticas a que estão expostos. Existem ainda outros entraves no projeto, uma vez que os sensores utilizados são de baixo custo e necessitam de ser calibrados. Como resposta a este problema, os investigadores desenvolveram uma estratégia que faz com que os sensores sejam calibrados automaticamente, periodicamente. A calibração é feita totalmente na infraestrutura *backend*, o que permite que o *hardware* e o *software* dos DCUs não sofram qualquer tipo de alteração.

5 Plano de Gestão de Dados para o projeto UrbanSense

No desenvolvimento de projetos de investigação, os investigadores preocupam-se com várias questões relativas à proveniência, à partilha, ao acesso e à preservação dos dados, nomeadamente, características dos dados, tipologias, disponibilidade para outros investigadores, monitorização dos acessos e método de armazenamento, entre outras.

A existência de um Plano de Gestão de Dados (PGD) num projeto de investigação permite que a gestão dos dados seja feita de forma eficaz ao longo do ciclo de vida dos dados, o que resulta em investigações bem-sucedidas.

Como já referido anteriormente, com o aumento da consciencialização da importância da elaboração de um PGD, os financiadores de projetos de investigação têm cada vez mais exigido o PGD a acompanharem os projetos de investigação.

Assim sendo, e uma vez que o projeto UrbanSense não continha um PGD sentiu-se a necessidade de criar um plano que acompanhasse e orientasse a gestão dos dados UrbanSense, já que este permite descrever quais os dados que são criados ou recolhidos, armazenados, geridos e partilhados ao longo do projeto e consciencializar os investigadores e todos os envolvidos no projeto para assegurarem o cumprimento de boas práticas relativamente aos dados.

A adoção de ferramentas que auxiliam na elaboração do PGD serve de guia durante o desenvolvimento do plano. Aquando da elaboração do PGD, apesar da existência de outras plataformas e modelos que auxiliam na criação de um PGD, optou-se pela utilização da DMPonline, criada pelo DCC (Digital Curation Center), dado que esta é uma ferramenta online que permite criar um PGD de acordo com as necessidades do projeto ao qual será aplicado e dos requisitos dos financiadores do projeto. Esta ferramenta é composta por algumas questões que orientam o desenvolvimento do PGD, como se pode ver na Figura 9, e ainda permite partilhá-lo com outros utilizadores e exportá-lo em vários formatos, o que possibilita que sejam feitas alterações mesmo depois de fazer o *download* do PGD e sejam acrescentadas outras questões para além das existentes na plataforma.

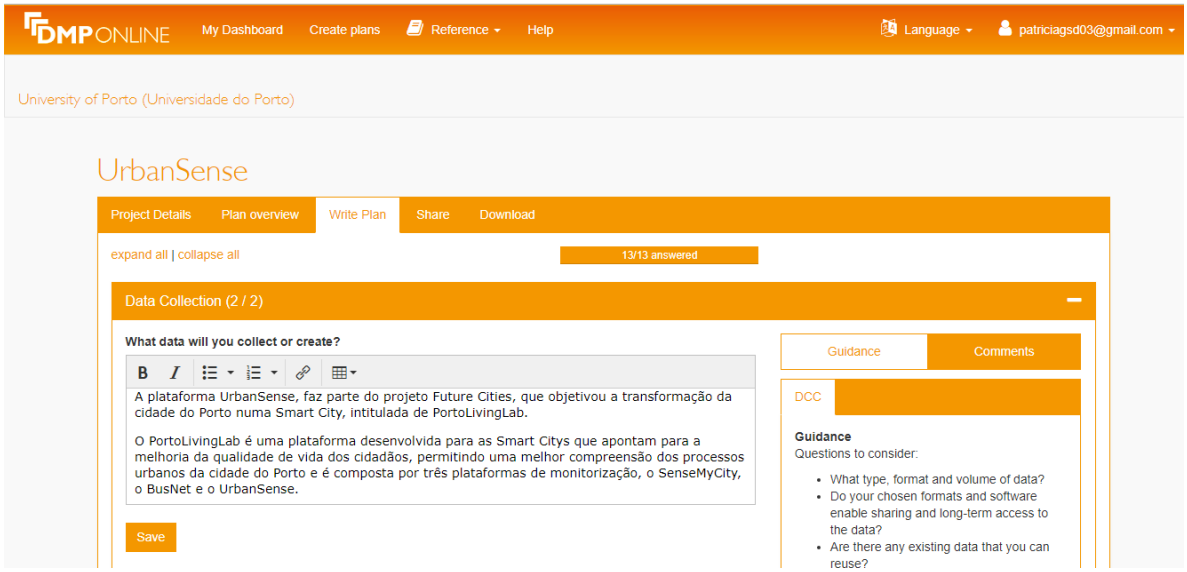


Figura 9 Plataforma DMPonline.

Fonte: DMPonline

Disponível em: https://dmponline.dcc.ac.uk/

O PGD desenvolvido para o projeto UrbanSense, disponível para análise mais detalhada no anexo I, aborda vários requisitos relativos à recolha de dados, à documentação e metadados, ética e conformidade legal, armazenamento e *backup*, seleção e preservação, partilha de dados e responsabilidades e recursos. Num primeiro ponto explicita-se que dados são recolhidos através das várias unidades de recolha de dados, onde estão localizadas as várias estações e respetivos sensores, no caso ambientais, e de que forma os dados são recolhidos. Relativamente à documentação e metadados que acompanham os dados, enumeram-se os modelos de metadados que são utilizados para descrever os dados através da plataforma de gestão de dados, Dendro, assim como os descritores selecionados. No que respeita às questões éticas e legais, no caso do projeto UrbanSense, não há problemas éticos relacionados com a recolha dos dados e qualquer publicação relacionada com o projeto estará dependente de autorização do responsável pelo projeto, uma vez que que o mesmo está a ser realizado em colaboração com vários parceiros, também especificados no PGD. Relativamente à preservação e partilha dos dados, no caso dos dados UrbanSense que representam dados observacionais recolhidos num determinado momento e local e que por essa razão não podem ser gerados de novo, leva a que estes sejam considerados de valor a longo prazo e por isso serão mantidos e preservados num repositório de dados de investigação de acesso aberto – INESC TEC.

Na parte final do plano de gestão de dados é referida a responsabilidade por parte do INESC TEC na organização e armazenamento dos dados e pela gestão do repositório e são ainda referidos os vários recursos necessários para o desenvolvimento do plano.

6 Os dados de investigação da Plataforma UrbanSense

Segundo a Curtin University (2018)[26], os "dados de investigação são quaisquer materiais físicos e/ou digitais que são recolhidos, observados ou criados em atividades de investigação para fins de análise para produzir resultados de investigação originais ou trabalhos criativos".

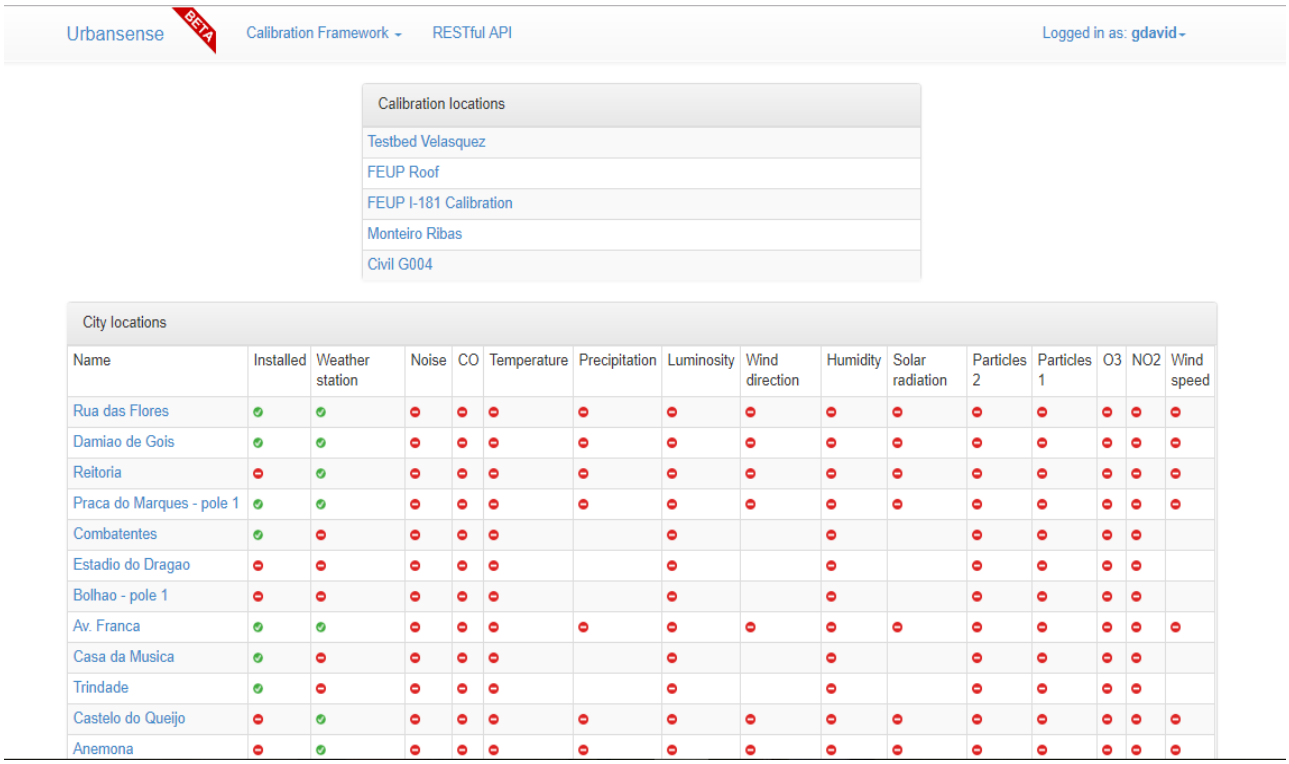
Os dados de investigação podem ser recolhidos/produzidos com diferentes finalidades e através de vários processos e por isso podem ser categorizados de acordo com a sua tipologia. Podemos falar de dados brutos, sem qualquer tipo de processamento ou dados processados que sofreram algum tipo de análise e ainda dados experimentais, observacionais ou simulacionais.

Partindo deste princípio os dados de investigação referentes ao projeto UrbanSense surgem no contexto do domínio ambiental e objetivam compreender os processos urbanos da cidade do Porto. Os dados resultantes das investigações são dados observacionais, ou seja, dados recolhidos em tempo real pelos vários sensores espalhados pela cidade, que estão dependentes de um determinado local e momento, não podendo ser reproduzidos novamente, como é o caso da temperatura registada no dia 25 de agosto de 2017 pelas 23:42:41 horas. Assim é fundamental a análise e tratamento dos dados observacionais para que sejam partilhados e preservados a longo prazo.

Os dados analisados neste projeto, não sofrem qualquer tipo de processamento, tratando-se de dados brutos.

Com o objetivo de explorar e analisar os dados do projeto, foi necessário aceder à plataforma que albergava os dados (ver Figura 10) através das credencias disponibilizadas pelos responsáveis do projeto.

[26] Curtin University: http://libguides.library.curtin.edu.au/c.php?g=202401&p=1333152



Name	Installed	Weather station	Noise	CO	Temperature	Precipitation	Luminosity	Wind direction	Humidity	Solar radiation	Particles 2	Particles 1	O3	NO2	Wind speed
Rua das Flores	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Damiao de Gois	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Reitoria	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Praca do Marques - pole 1	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Combatentes	✓	✗	✗	✗	✗		✗		✗		✗	✗	✗	✗	
Estadio do Dragao	✗	✗	✗	✗	✗		✗		✗		✗	✗	✗	✗	
Bolhao - pole 1	✗	✗	✗	✗	✗		✗		✗		✗	✗	✗	✗	
Av. Franca	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Casa da Musica	✓	✗	✗	✗	✗		✗		✗		✗	✗	✗	✗	
Trindade	✓	✗	✗	✗	✗		✗		✗		✗	✗	✗	✗	
Castelo do Queijo	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Anemona	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗

Figura 10 Plataforma UrbanSense.

Fonte: UrbanSense

Disponível em: http://cloud.futurecities.up.pt/~urbansenserpi/UrbansenseSite/

(necessita de autenticação)

A Plataforma, compreende dados recolhidos por 13 tipos de sensores, desde 22 de julho de 2014 a 26 de agosto de 2017, que se encontram dispostos em 23 locais da cidade do Porto.

Para ter acesso aos dados UrbanSense, foi necessário fazer o *download* dos dados utilizando para isso a plataforma (ver Figura 11).

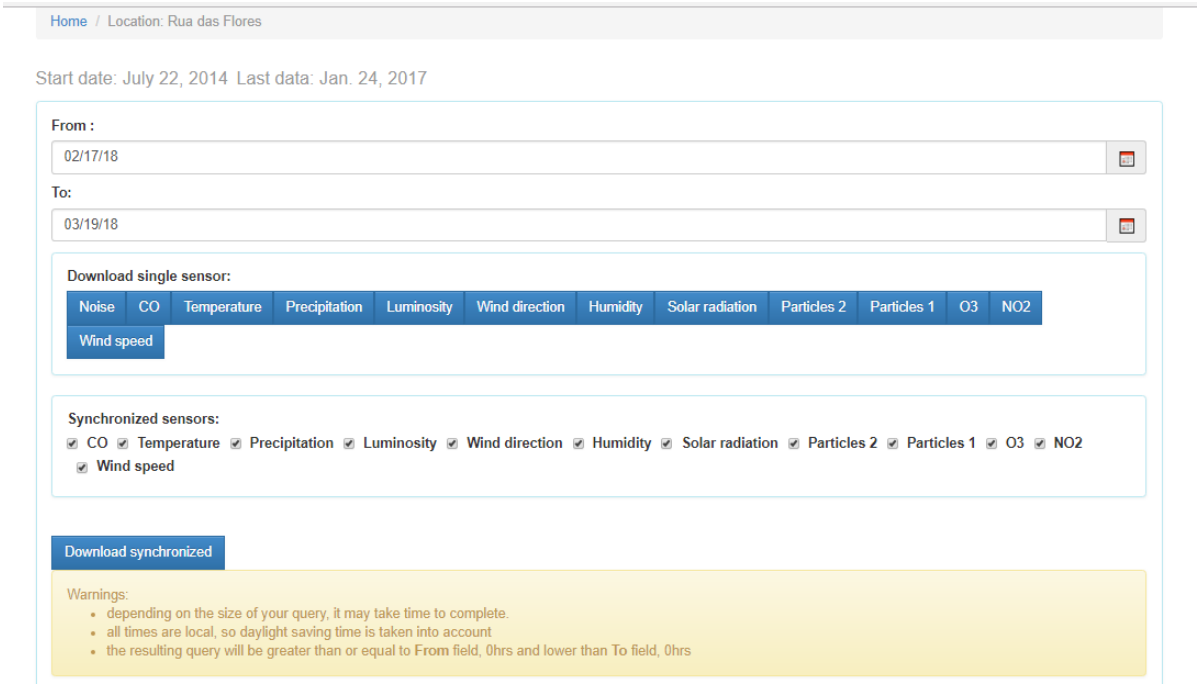
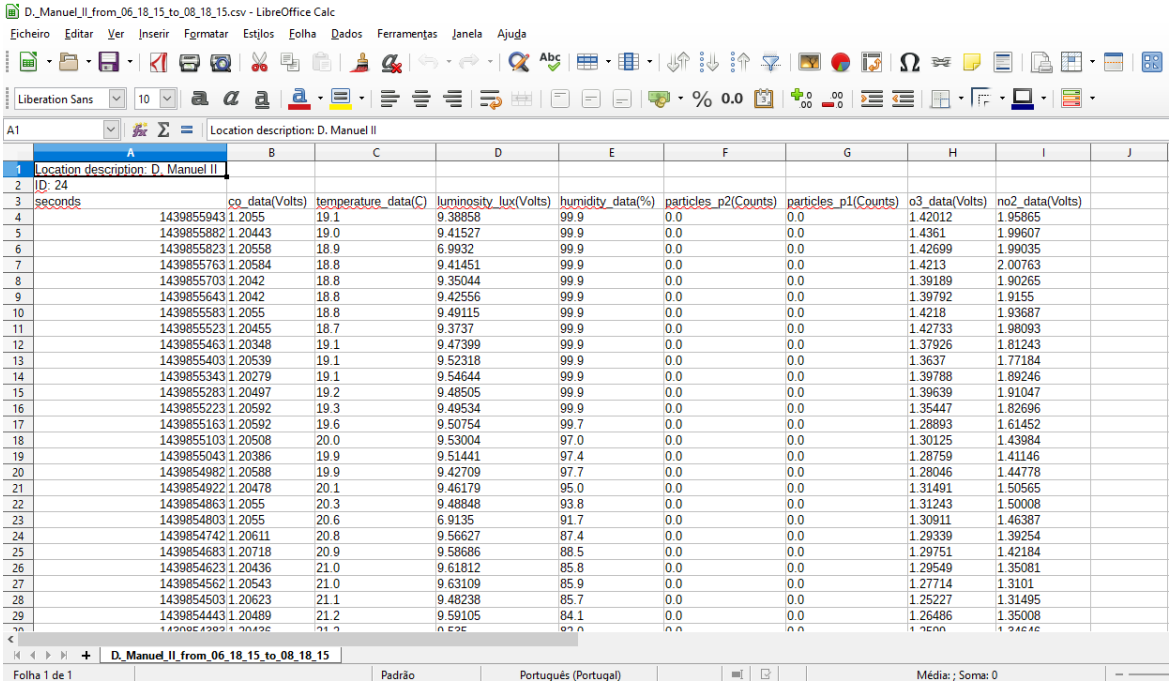


Figura 11 Local para a realização do *download.*

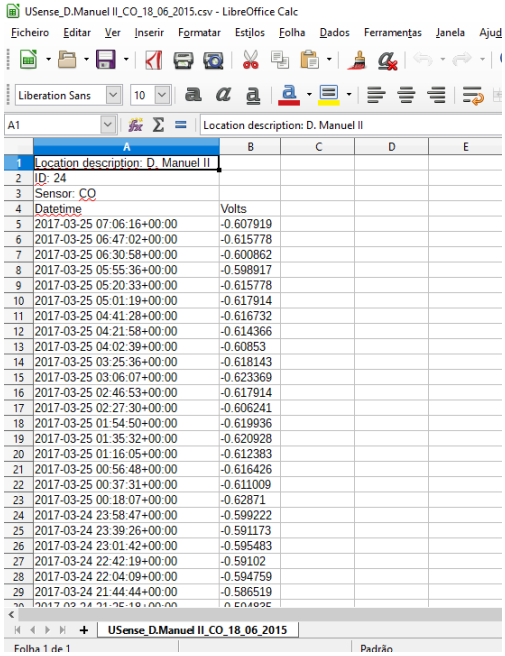
Fonte: UrbanSense

Inicialmente tentou-se fazer o *download* sincronizado de todos os sensores de determinada estação. No entanto, verificou-se que a informação obtida não estava completa uma vez que ao fazer o *download* sincronizado a informação que compunha o excel, para além das medições realizadas pelos vários sensores, correspondia apenas aos segundos (ver Figura 12), enquanto que ao realizar o *download* individualmente, ou seja, para cada estação, sensor a sensor, a informação obtida é mais completa, pois como se pode ver na Figura 13, para além da medição realizada pelo sensor, é apresentada a data e a hora em que cada medição foi registada.



	A	B	C	D	E	F	G	H	I
1	Location description: D. Manuel II								
2	ID: 24								
3	seconds	co_data(Volts)	temperature_data(C)	luminosity_lux(Volts)	humidity_data(%)	particles_p2(Counts)	particles_p1(Counts)	o3_data(Volts)	no2_data(Volts)
4	1439855943	1.2055	19.1	9.38858	99.9	0.0	0.0	1.42012	1.95865
5	1439855882	1.20443	19.0	9.41527	99.9	0.0	0.0	1.4361	1.99607
6	1439855823	1.20558	18.9	6.9932	99.9	0.0	0.0	1.42699	1.99035
7	1439855763	1.20584	18.8	9.41451	99.9	0.0	0.0	1.4213	2.00763
8	1439855703	1.2042	18.8	9.35044	99.9	0.0	0.0	1.39189	1.90265
9	1439855643	1.2042	18.8	9.42556	99.9	0.0	0.0	1.39792	1.9155
10	1439855583	1.2055	18.8	9.49115	99.9	0.0	0.0	1.4218	1.93687
11	1439855523	1.20455	18.7	9.3737	99.9	0.0	0.0	1.42733	1.98093
12	1439855463	1.20348	19.1	9.47399	99.9	0.0	0.0	1.37926	1.81243
13	1439855403	1.20539	19.1	9.52318	99.9	0.0	0.0	1.3637	1.77184
14	1439855343	1.20279	19.1	9.54644	99.9	0.0	0.0	1.39788	1.89246
15	1439855283	1.20497	19.2	9.48505	99.9	0.0	0.0	1.39639	1.91047
16	1439855223	1.20592	19.3	9.49534	99.9	0.0	0.0	1.35447	1.82696
17	1439855163	1.20592	19.6	9.50754	99.7	0.0	0.0	1.28893	1.61452
18	1439855103	1.20508	20.0	9.53004	97.0	0.0	0.0	1.30125	1.43984
19	1439855043	1.20386	19.9	9.51441	97.4	0.0	0.0	1.28759	1.41146
20	1439854982	1.20588	19.9	9.42709	97.7	0.0	0.0	1.28046	1.44778
21	1439854922	1.20478	20.1	9.46179	95.0	0.0	0.0	1.31491	1.50565
22	1439854863	1.2055	20.3	9.48848	93.8	0.0	0.0	1.31243	1.50008
23	1439854803	1.2055	20.6	6.9135	91.7	0.0	0.0	1.30911	1.46387
24	1439854742	1.20611	20.8	9.56627	87.4	0.0	0.0	1.29339	1.39254
25	1439854683	1.20718	20.9	9.58686	88.5	0.0	0.0	1.29751	1.42184
26	1439854623	1.20436	21.0	9.61812	85.8	0.0	0.0	1.29549	1.35081
27	1439854562	1.20543	21.0	9.63109	85.9	0.0	0.0	1.27714	1.3101
28	1439854503	1.20623	21.1	9.48238	85.7	0.0	0.0	1.25227	1.31495
29	1439854443	1.20489	21.2	9.59105	84.1	0.0	0.0	1.26486	1.35008

Figura 12 Exemplo de ficheiro obtido através do download sincronizado.



	A	B	C	D	E
1	Location description: D. Manuel II				
2	ID: 24				
3	Sensor: CO				
4	Datetime	Volts			
5	2017-03-25 07:06:16+00:00	-0.607919			
6	2017-03-25 06:47:02+00:00	-0.615778			
7	2017-03-25 06:30:58+00:00	-0.600862			
8	2017-03-25 05:55:36+00:00	-0.598917			
9	2017-03-25 05:20:33+00:00	-0.615778			
10	2017-03-25 05:01:19+00:00	-0.617914			
11	2017-03-25 04:41:28+00:00	-0.616732			
12	2017-03-25 04:21:58+00:00	-0.614366			
13	2017-03-25 04:02:39+00:00	-0.60853			
14	2017-03-25 03:25:36+00:00	-0.618143			
15	2017-03-25 03:06:07+00:00	-0.623369			
16	2017-03-25 02:46:53+00:00	-0.617914			
17	2017-03-25 02:27:30+00:00	-0.606241			
18	2017-03-25 01:54:50+00:00	-0.619936			
19	2017-03-25 01:35:32+00:00	-0.620928			
20	2017-03-25 01:16:05+00:00	-0.612383			
21	2017-03-25 00:56:48+00:00	-0.616426			
22	2017-03-25 00:37:31+00:00	-0.611009			
23	2017-03-25 00:18:07+00:00	-0.62871			
24	2017-03-24 23:58:47+00:00	-0.599222			
25	2017-03-24 23:39:26+00:00	-0.591173			
26	2017-03-24 23:01:42+00:00	-0.595483			
27	2017-03-24 22:42:19+00:00	-0.59102			
28	2017-03-24 22:04:09+00:00	-0.594759			
29	2017-03-24 21:44:44+00:00	-0.586519			

Figura 13 Exemplo de ficheiro obtido através do download individual.

Por este motivo optou-se por fazer o *download* individualmente, o que implicou um maior esforço e tempo pois trata-se de uma grande quantidade de dados, o que torna o processo mais lento e que originou outros problemas identificados numa fase seguinte, aquando das entrevistas com os investigadores/colaboradores. Para fazer o download individual inseriu-se a data do primeiro dado recolhido naquela estação e a última data em que o sensor recolheu informação (a informação da data estava disponível na plataforma dos dados UrbanSense),

selecionando-se de seguida o sensor do qual se pretendia obter os dados. Em alguns casos, ao fazer o download, o ficheiro resultante (em formato *.csv), ultrapassava a capacidade máxima da aplicação do software utilizado para abrir o ficheiro, pelo que foi necessário dividir os dados que esse ficheiro continha, em vários outros ficheiros, resultando assim num total de 328 ficheiros obtidos a partir da plataforma UrbanSense.

Ainda nesta fase, optou-se por organizar os ficheiros (ver Figura 14) numa pasta principal, composta por 23 subpastas que correspondiam às 23 estações, cada subpasta continha ainda outras pastas, de acordo com os sensores existentes naquela estação, sendo que cada pasta continha o ficheiro correspondente ao sensor e à estação e ainda nos casos em que foi necessário dividir o ficheiro *.csv em vários outros ficheiros, como referido anteriormente, a pasta continha vários ficheiros. Ainda na fase de organização dos dados, com o objetivo de normalizar os *data sets* foi criado um código de referência composto pela abreviatura do nome do projeto "USense", seguido da estação a que os dados daquele documento correspondem e do sensor e por fim a data de início de recolha dos dados do sensor a que se refere, resultando por exemplo em "USense_24_de_Agosto_CO_25_06_2015".

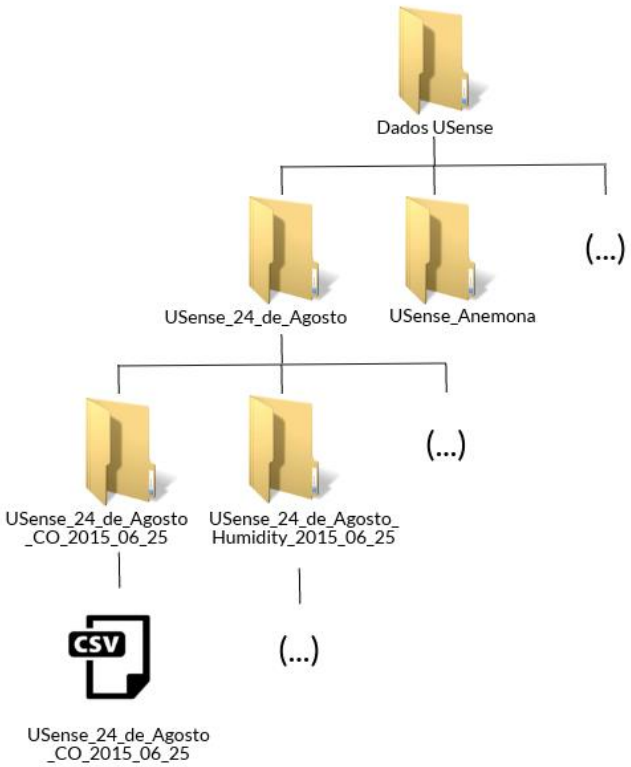


Figura 14 Esquema representativo de organização dos *datasets*.

6.1 Análise e inventariação dos dados

A análise e inventariação dos dados representou uma parte importante do processo de gestão de dados de investigação, como afirma a University of Minnesota (2018)[27] "uma componente importante do Research Data Management (RDM) é realizar um inventário dos dados, um *brainstorm* de todos os dados que serão recolhidos" uma vez que o documento fornece uma visão geral dos dados e permite registar detalhadamente os dados recolhidos no projeto UrbanSense e compreender a dimensão dos mesmos.

A elaboração de um inventário dos dados do projeto pode representar um documento útil no que respeita à utilização dos dados, quer para os investigadores do projeto, quer para outros que os venham a utilizar. Este registo trará ainda outros benefícios futuros pois permitirá que em caso de perda de algum dos dados, estes sejam facilmente identificáveis. Os autores Horton et al. (2011) no artigo *Data Management Recommendations for Research Centres and Programmes*, refere três aspetos em que a criação de um inventário de dados pode ajudar, nomeadamente no que respeita à divulgação dos dados, à sensibilização dos recursos de dados locais e à gestão de dados durante o processo de investigação.

O inventário elaborado para o projeto UrbanSense, que pode ser consultado mais detalhadamente no anexo II, teve como foco principal dar uma visão geral dos dados recolhidos e foi elaborado com base na organização dos dados realizada na fase anterior. Neste inventário optou-se por registar o nome do *data set* que corresponde ao código de referência anteriormente criado, a estação a que o mesmo estava associado, a data do primeiro e do último registo recolhido por aquele sensor, o tipo de dados que o *data set* continha, que neste caso são designados de dados observacionais e ainda o tamanho dos ficheiros e o número de registos existentes em cada *data set* (ver Figura 15).

Inventário dados UrbanSense							
ID	Nome do *dataset*	Estação	Data		Tipo de dados	Tamanho do ficheiro	Nº registos
			Início	Fim			
1	USense_24_de_Agosto_CO_2015_06_25	24 de Agosto	25-06-2015	04-02-2017	Dados observacionais (amostra)	2,78 MB	83491
2	USense_24_de_Agosto_Humidity_2015_06_25	24 de Agosto	25-06-2015	26-08-2017	Dados observacionais (amostra)	7,61 MB	252766
3	USense_24_de_Agosto_Luminosity_2015_06_25	24 de Agosto	25-06-2015	03-02-2017	Dados observacionais (amostra)	2,70 MB	81169
4	USense_24_de_Agosto_NO2_2015_06_25	24 de Agosto	25-06-2015	05-01-2017	Dados observacionais (amostra)	2,33 MB	70155
5	USense_24_de_Agosto_Noise_2015_07_29	24 de Agosto	29-07-2015	26-08-2017	Dados observacionais (amostra)	31 MB	1028687
6	USense_24_de_Agosto_O3_2015_06_25	24 de Agosto	25-06-2015	04-01-2017	Dados observacionais (amostra)	2,33 MB	70025
7	USense_24_de_Agosto_Particles1_2015_06_25	24 de Agosto	25-06-2015	26-08-2017	Dados observacionais (amostra)	8,34 MB	275641
8	USense_24_de_Agosto_Particles2_2015_06_25	24 de Agosto	25-06-2015	26-08-2017	Dados observacionais (amostra)	8,33 MB	275641
9	USense_24_de_Agosto_Precipitation_2015_06_25	24 de Agosto	25-06-2015	24-03-2017	Dados observacionais (amostra)	4,43 MB	149782
10	USense_24_de_Agosto_Solar radiation_2015_06_25	24 de Agosto	25-06-2015	04-02-2017	Dados observacionais (amostra)	4,01 MB	117179
11	USense_24_de_Agosto_Temperature_2015_06_25	24 de Agosto	25-06-2015	26-08-2017	Dados observacionais (amostra)	8,38 MB	275318
12	USense_24_de_Agosto_Wind direction_2015_06_25	24 de Agosto	25-06-2017	24-03-2017	Dados observacionais (amostra)	3,29 MB	112238
13	USense_24_de_Agosto_Wind speed_2015_06_25	24 de Agosto	25-06-2017	24-03-2017	Dados observacionais (amostra)	4,69 MB	158765
14	USense_Anemona_CO_2015_05_24	Anemona	24-05-2015	02-08-2015	Dados observacionais (amostra)	793 KB	23278
15	USense_Anemona_Humidity_2015_05_24	Anemona	24-05-2015	16-08-2015	Dados observacionais (amostra)	3,42 MB	112389
16	USense_Anemona_Luminosity_2015_05_24	Anemona	24-05-2015	29-07-2015	Dados observacionais (amostra)	2,84 MB	84159

Figura 15 Excerto do inventário dos dados UrbanSense.

[27] University of Minnesota: http://libguides.d.umn.edu/c.php?g=307731&p=2409893

Parte destas componentes, selecionadas para constituir o inventário, já estavam presentes nos *data sets* recolhidos através da plataforma UrbanSense, no entanto as restantes implicaram uma análise detalhada dos vários *data sets*. Para que fosse possível identificar a data de início da recolha de dados por cada sensor, assim como a data final foi necessário analisar cada *data set* individualmente. Apesar da plataforma UrbanSense identificar a data de início e de fim dos dados recolhidos em cada estação, verificou-se que essas datas não se aplicavam a todos os sensores que compõem essa estação, por exemplo no Campo 24 de Agosto, enquanto que o sensor que mede a luminosidade começou a registar dados a partir do dia 25 de junho de 2015 até ao dia 05 de janeiro de 2017, o sensor referente ao ruído, só começou a registar dados a partir do dia 29 de julho de 2015 até 26 de agosto de 2017.

Relativamente ao tipo de dados, desde logo se identificaram como dados observacionais, tratando-se de dados precisos que foram recolhidos num momento específico, não podendo ser recolhidos novamente. Mais tarde, aquando da realização de entrevistas com alguns dos envolvidos, quando lhes foi questionado acerca das características dos dados, particularmente da tipologia dos dados, foi referido que para eles são considerados amostras, pelo que se optou por qualificar no inventário como "Dados observacionais (amostra)", de forma a não se distanciar dos termos utilizados pelos investigadores/colaboradores do projeto.

No que respeita ao tamanho do ficheiro e ao número de registos, estes foram dos elementos mais importantes do inventário dado que, em caso de perda dos dados, estes elementos permitirão que essa falha seja facilmente identificável.

Esta fase de inventariação dos dados, permitiu confirmar a dimensão dos dados, uma vez que foram inventariados um total de 328 ficheiros, compostos por 101 616 023 registos. Ainda nesta fase, com o objetivo de entender melhor o fluxo de dados produzidos durante o projeto de investigação procedeu-se à elaboração de uma entrevista com algumas das pessoas envolvidas no projeto, nomeadamente um técnico e duas investigadoras do projeto.

6.1.1 Análise das entrevistas aos investigadores

A entrevista é um instrumento de recolha de dados muito utilizado para uma primeira interpretação do estudo pois permite que o entrevistador obtenha informações mais específicas acerca do projeto e ainda que o entrevistado tenha uma perceção mais extensiva sobre gestão de dados e do que se pretende realizar.

O recurso às entrevistas constituiu um acréscimo de valor para o trabalho desenvolvido nesta dissertação, uma vez que permitiu verificar de que forma decorreu o fluxo de dados recolhidos durante o projeto e contribuiu para numa fase posterior selecionar os metadados que compõem as ontologias criadas e os metadados utilizados para descrever os *data sets* do projeto. O contacto com os investigadores/colaboradores revelou-se também importante na

medida em que permitiu identificar algumas falhas no *download* dos dados através da plataforma UrbanSense entre outras questões que limitaram este estudo.

Para a elaboração da entrevista foi desenvolvido um guião, que se pode ver no anexo III, "Guião de Entrevista aos Investigadores do Projeto UrbanSense", que foi previamente enviado aos investigadores/colaboradores para que tivessem a oportunidade de se preparar para a mesma. Este guião foi desenvolvido seguindo as orientações do *Data Curation Profile Toolkit*[28] e aborda questões sobre o ciclo de vida dos dados, a avaliação dos dados, a descrição dos dados, visualização, controlo de acesso e impacto/resultados da utilização dos dados.

Assim, quando foi pedido aos entrevistados para falarem um pouco do que acontece durante o processo de recolha de dados, os mesmos afirmaram que durante o processo poderiam recolher diferentes tipos de informação, dependendo do tipo de equipamento que utilizam o que consequentemente leva a que a qualidade e a granularidade dessa informação seja diferente. Os mesmos disseram que em muitos casos poderia mesmo haver situações de risco de intromissão na vida das pessoas, no entanto isso não se verificava no caso dos equipamentos, nomeadamente do sensor do ruído, que estão espalhados pela cidade do Porto, uma vez que estes recolhem níveis sonoros e não o sinal acústico. Foi ainda durante a entrevista, que foi colocada a questão acerca de uma dúvida que persistia relativa ao sensor do ruído e do pré-processamento que o mesmo sofria, o que foi esclarecido que o sensor recolhe os dados numa determinada granularidade, entenda-se por granularidade o nível de detalhamento dos dados, e que pode ser pedido para que os sensores recolham dados numa granularidade diferente.

Quando se recolhe dados existem vários aspetos a ter em conta, nomeadamente no que respeita à qualidade dos dados recolhidos pois esta pode ser influenciada por vários fatores. Foram então identificados pelos investigadores entrevistados os seguintes fatores que podem influenciar os dados:

- Qualidade do equipamento;
- Montagem do equipamento;
- Método de recolha dos dados (influencia sobretudo a quantidade dos dados);

Foram ainda referidos outros aspetos que não influenciam diretamente a qualidade dos dados, mas podem influenciar os dados, como é o caso do ambiente em que os sensores estão inseridos. Foram também referidas algumas situações específicas que influenciaram a recolha dos dados no projeto UrbanSense, como é o caso das ligações à internet, que falhavam muitas vezes e por isso durante alguns períodos não era possível recolher dados, e do suporte, como o

[28] Mais informações em: Data Curation Profile Toolkit http://datacurationprofiles.org/about

projeto estava vinculado à Câmara Municipal do Porto, a resolução de alguns problemas dependia da instituição e não dos próprios investigadores.

Quando contactados para a realização das entrevistas, os investigadores foram esclarecidos sobre o objetivo principal da realização das mesmas. Se por um lado se pretendia compreender o processo de recolha dos dados por outro, pretendia-se sobretudo perceber quais os dados mais importantes a analisar, disponibilizar e preservar, ou seja, estabelecer junto dos investigadores especialistas em cada domínio, parâmetros que seriam utilizados para distinguir dados "válidos" (dados que foram recolhidos enquanto os sensores estavam calibrados e com as condições propícias para a recolha) dos dados "inválidos". Essa avaliação dos dados iria permitir que numa fase seguinte fossem apenas disponibilizados no repositório do INESC TEC os dados válidos, que cumprissem os parâmetros estabelecidos pelos investigadores.

Contudo, os investigadores afirmaram que existem formas diferentes de lidar com os dados, se por exemplo os dados de investigação forem entregues aos investigadores em bruto, o seu tratamento deve ser feito da maneira que o investigador que os recebe e que os vai utilizar entende e como lhe for mais conveniente, tendo sempre em atenção que existem circunstâncias que estiveram associadas ao momento da medição, umas vezes identificadas outras vezes não.

Uma vez que a plataforma UrbanSense disponibiliza apenas dados brutos e que estes dados já foram utilizados noutras investigações, foi questionado aos investigadores se alguma vez fizeram algum tipo de análise para os dados, pelo que foi dito que nos trabalhos a decorrer fazem essa análise utilizando os dados brutos e aplicando-lhes o seu juízo de valor, ou seja, sugerem que cada vez que um investigador utilize o conjunto de dados, deve ser o próprio a estabelecer os parâmetros para os analisar. Apesar de ser esperado que dentro de um mesmo domínio os investigadores estabeleçam parâmetros semelhantes, pode, no entanto, esta análise resultar em alguns desvios uma vez que é aplicado um juízo de valor, que irá variar dependendo do investigador.

Posto isto, percebeu-se que este poderia ser um entrave para a realização de uma das tarefas previstas neste projeto de dissertação, a avaliação da qualidade dos dados, uma vez que segundo os entrevistados, os parâmetros que poderiam ser utilizados para fazer a avaliação da qualidade dos dados, devem ser estabelecidos pelos investigadores que os vão utilizar, de acordo com a finalidade dessa utilização.

Na tentativa de perceber se os investigadores estavam acostumados a utilizar algum tipo de documentação, metadados para descrever os dados foi-lhes questionado se têm por hábito fazer anotações quando estão a recolher ou analisar os dados. Sendo que os mesmos responderam que não sentiam necessidade de fazer anotações, porque os metadados que necessitam já estão incorporados na própria informação. Segundo eles, tanto no caso particular da poluição atmosférica como sonora e mesmo no clima, como são dados localizados

espacialmente e temporalmente, a informação de que necessitam já está anotada. Para eles, a informação de que necessitam (local, tempo e o tipo de poluente) já consta na plataforma UrbanSense, já está incorporada nos ficheiros que contêm os dados (ver Figura 16) e que por isso não fazia sentido estarem a introduzir duplicação da informação. Mesmo nos casos em que por exemplo se identificassem concentrações anormais durante um período de tempo, os próprios investigadores do domínio iriam procurar as possíveis causas para a ocorrência desse acontecimento, mesmo não tendo qualquer tipo de anotações junto dos dados e os investigadores não estarem envolvidos no projeto, facilmente interpretavam essas anormalidades.

Location description: 24 de Agosto Ferreira Cardoso	
ID: 9	
Sensor: CO	
Datetime	Volts
2017-02-04 13:06:57+00:00	1.03487
2017-02-04 12:33:51+00:00	0.532006
2017-02-04 12:09:54+00:00	0.179942
2017-02-04 11:09:57+00:00	0.034333
2017-02-04 10:49:07+00:00	0.013123

Figura 16 Excerto do data set com os metadados.

Fonte: Plataforma UrbanSense

Se ponderarmos um caso de um projeto que já terminou e e durante o qual não foi associada qualquer descrição aos dados desse projeto e anos mais tarde se voltam a utilizar esses dados, certamente há pormenores que irão falhar. É assim que surge a gestão de dados como forma de facilitar a reutilização dos dados, uma vez que permite documentar o processo desde o início até ao seu término. Partindo deste ponto de vista, os investigadores concordaram em parte que a informação adicional seria mais útil para terceiros do que para o próprio investigador, uma vez que segundo eles o próprio investigador sabe o que aconteceu, porque é que os dados foram recolhidos e como foi feita essa recolha, logo essa informação não seria necessária.

Durante a entrevista foi mostrado aos investigadores, o inventário dos dados UrbanSense realizado, para que estes expressassem a sua opinião acerca da informação que compunha o inventário e do código de referência utilizado nos vários *data sets*. Foi nesta fase que os investigadores identificaram anormalidades na quantidade de dados pois a plataforma não estava a devolver os dados na sua totalidade, facto que já tinha acontecido noutros momentos com os próprios investigadores. O ruído era dos sensores que mais falhas apresentava, pois não estavam de todo completos os dados e por isso foi decidido que teria de ser feito de novo o

download dos dados em falta, mas optando-se por selecionar períodos de tempo mais curtos, de forma a contornar a situação.

Uma vez que o objetivo final deste trabalho é a publicação dos dados num repositório para que estes estejam acessíveis para outros investigadores, questões como a visualização dos dados e o controlo de acesso devem ser levadas em conta. Concluiu-se que na perspetiva dos investigadores a forma como os dados deveriam estar disponíveis varia de acordo com a finalidade e o objetivo que os investigadores querem dar, assim como da pessoa que os vai utilizar. No caso das pessoas que vão utilizar os dados serem investigadores, estes vão tratar os dados da forma que mais lhe convém, no caso de ser o público em geral a utilizar os dados, estes devem ser analisados antes de serem disponibilizados.

Por outro lado, na perspetiva dos técnicos, deve-se optar por uma representação mais básica possível, o que presume o mínimo possível de processamento dos dados.

Relativamente ao impacto que a partilha de dados poderia ter com terceiros, concluiu-se que a divulgação destes dados é importante, pois pode gerar outro tipo de informação de mais fácil interpretação para quem os vai receber, uma vez que estes dados podem ser aplicados e utilizados para vários tipos de atividades na cidade do Porto.

Contudo, ao longo das entrevistas os investigadores demonstraram alguma indecisão relativamente à partilha e reutilização dos dados. Por um lado, os investigadores afirmam que os dados poderão não ter muito interesse para outros investigadores, uma vez que a informação que será disponibilizada corresponde a um período de tempo passado. Por outro lado, afirmam que os dados poderão servir para medir tendências, como é o caso específico da qualidade do ar que pode servir para fazer modelos de previsão. Aquilo que se identifica é que estes dados poderão ser utilizados em diversos âmbitos e podem sobretudo ser utilizados para prever, perceber tendências, quer ao nível do clima, meio ambiente, possíveis anomalias que exigem intervenções urbanas, planeamento urbano, entre outros.

6.2 Análise e seleção de modelo de metadados para smart cities

Um modelo de metadados é composto por metadados que descrevem um conjunto de dados de investigação que representam as informações específicas necessárias para contextualizar os dados.

O conjunto de dados analisado no contexto deste trabalho, está inserido no domínio ambiental e faz parte de um projeto direcionado para as *smart cities*, composta por três projetos: BusNet, SenseMyCity e UrbanSense.

Ao invés de o foco ser apenas o projeto individual UrbanSense, tentou-se pensar nos três projetos (BusNet, SenseMyCity e UrbanSense) como um todo. Numa perspetiva futura, em que os dados dos outros dois projetos possam vir a ser descritos, optou-se por analisar e

selecionar modelos de metadados desenvolvidos para o domínio da IoT e para a descrição de sensores, para que ficassem preparados no Dendro, modelos de metadados que pudessem vir a ser utilizados para a descrição futura dos outros dados. O objetivo passou pela reutilização de metadados já existentes nestes modelos, mas que ainda não estavam incorporados na plataforma de gestão de dados, Dendro.

Os modelos de metadados selecionados para este estudo têm como objetivo a descrição de dados relacionados com as *smart cities*. Para proceder à criação do modelo, foram realizadas entrevistas, analisadas no capítulo 6.1.1, entre os envolvidos no processo de descrição dos dados (investigadores e o curador de dados). Esta entrevista teve como principal objetivo adquirir conhecimento na área, nomeadamente ao nível dos instrumentos e métodos de recolha de dados, armazenamento, publicação, entre outros.

Durante as entrevistas os investigadores responderam a perguntas mais gerais como o caso da descrição do processo de investigação, preservação e partilha dos dados, de forma a compreender as suas opiniões e ainda responderam a perguntas mais específicas acerca dos procedimentos durante a investigação.

Após a realização das entrevistas e também da leitura de artigos relacionados com o projeto PortoLivingLab, analisaram-se alguns modelos de metadados do domínio, entre os quais, Ontology for Meteorological Sensores (aws), M3-lite, Climate and Forecast (CF) e Semantic Sensor Network Ontology (SSN). Dentre os analisados selecionaram-se dois modelos de metadados o M3-lite e Semantic Sensor Network Ontology (SSN), que foram os que estavam mais relacionados com os projetos e os dados recolhidos nesses projetos. O M3-lite engloba vários domínios como a classificação de dispositivos, unidades de medida, diferentes tipos de sensores e o SSN é utilizado para descrever sensores, o tempo de implementação, precisão, capacidades entre outros.

Posteriormente, depois de selecionados os modelos de metadados, foi necessário analisar os descritores que compunham cada um dos modelos. Uma vez que não sendo um trabalho focado na seleção e análise de modelos de metadados para a área e o tempo foi dividido para a realização de outras tarefas, optou-se por selecionar um conjunto reduzido de descritores, específicos do domínio, mas com alguma generalidade, que possam ser utilizados para contextualizar de uma forma geral os projetos PortoLivingLab.

Ao todo foram selecionados 23 descritores, 10 fazem parte do modelo de metadados M3-lite e 13 do Semantic Sensor Network Ontology.

Na Tabela 3 é possível visualizar os descritores selecionados que fazem parte do modelo de metadados M3-lite assim como uma descrição de cada descritor, retirada do modelo de metadados original, que depois foi incorporada na plataforma Dendro.

Tabela 3 Descritores selecionados no modelo de metadados M3-lite.
Fonte: M3-lite Taxonomy

Disponível em: http://ontology.fiesta-iot.eu/ontologyDocs/fiesta-iot/doc

M3-lite	
Descritor	**Descrição**
Calibration measurement type	If the measurement was taken when the sensor was being calibrated.
Connected Users	Connected Users to a communication channel/platform.
Coordinates	Triples for location in the (Latitude, Longitude, Altitude) format.
Count	Number of available particular things.
Device	A device is a physical piece of technology - a system in a box. Devices may of course be built of smaller devices and software components (i.e. systems have components).
Distance	Measure of how far apart objects are.
Measurement type	Type of Measurement done using a device.
Transportation DOI	Transportation, Smart Car/Vehicle, Intelligent Transport System (ITS) as an Internet of Things (IoT) applicative domain.
Unit	A Unit is an abstract classifier that represents the concept of measurement unit that is defined as real scalar quantity, defined and adopted by convention, with which any other quantity of the same kind can be compared to express the ratio of the two quantities as a number.
Working State	Indicator of whether a person (or object) is working or not.

Na Tabela 4 encontram-se os descritores selecionados e as respetivas descrições, retiradas do modelo de metadados original, que integram o modelo de metadados SSN.

Tabela 4 Descritores selecionados no modelo de metadados SSN.
Fonte: Semantic Sensor Network Ontology

Disponível em: https://www.w3.org/2005/Incubator/ssn/ssnx/ssn.html

Semantic Sensor Network Ontology – SSN	
Descritor	**Descrição**
End Time	The exact moment when something ends.

Frequency	The smallest possible time between one observation and the next.
Measurement Range	The set of values that the sensor can return as the result of an observation under the defined conditions with the defined measurement properties. (If no conditions are specified or the conditions do not specify a range for the observed qualities, the measurement range is to be taken as the condition for the observed qualities.)
Observation	An Observation is a Situation in which a Sensing method has been used to estimate or calculate a value of a Property of a FeatureOfInterest. Links to Sensing and Sensor describe what made the Observation and how; links to Property and Feature detail what was sensed; the result is the output of a Sensor; other metadata details times etc.
Observation Result	Relation linking an Observation (i.e., a description of the context, the Situation, in which the observation was made) and a Result, which contains a value representing the value associated with the observed Property.
Observation Sampling Time	Rebadged as phenomenon time in [O&M]. The phenomenon time shall describe the time that the result applies to the property of the feature-of-interest. This is often the time of interaction by a sampling procedure or observation procedure with a real-world feature. The sampling time is the time that the result applies to the feature-of-interest. This is the time usually required for geospatial analysis of the result.
Plataform	An Entity to which other Entities can be attached - particularly Sensors and other Platforms. For example, a post might act as the Platform, a bouy might act as a Platform, or a fish might act as a Platform for an attached sensor.
Region	The region where the object is located or connected.
Sensing Method Used	A (measurement) procedure is a detailed description of a measurement according to one or more measurement principles and to a given measurement method, based on a measurement model and including any calculation to obtain a measurement result [VIM 2.6]
Sensor	A sensor can do (implements) sensing: that is, a sensor is any entity that can follow a sensing method and thus observe some Property of a FeatureOfInterest. Sensors may be physical devices,

	computational methods, a laboratory setup with a person following a method, or any other thing that can follow a Sensing Method to observe a Property.
Sensor Data Sheet	A data sheet records property of a sensor. A data sheet might describe for example the accuracy in various conditions, the power use, the types of connectors that the sensor has, etc. Generally, a sensor's properties are recorded directly (with hasMeasurementCapability, for example), but the data sheet can be used for example to record the manufacturers specifications verses observed capabilities, or if more is known than the manufacturer specifies, etc. The data sheet is an information object about the sensor's properties, rather than a direct link to the actual properties themselves.
Start Time	The exact moment when something started.
System	System is a unit of abstraction for pieces of infrastructure (and we largely care that they are) for sensing. A system has components, its subsystems, which are other systems.

Após a seleção dos descritores que iriam incorporar os modelos de metadados para as *smart cities*, foram enviados ao investigador responsável pelo projeto, para verificar se os descritores indicados eram suficientes para contextualizar os dados, para que estes sejam recuperados, interpretados e reutilizados por outros investigadores. Foi então enviado ao investigador um documento MSWord, com os descritores selecionados e as respetivas definições.

De um modo geral, o investigador demonstrou alguma divergência no que se refere às definições de alguns descritores, no entanto sendo que estes descritores foram reutilizados, as definições utilizadas para os explicar foram retiradas das ontologias originais pelo que se decidiu manter a definição. Em alguns casos, o investigador propôs a alteração do descritor, uma vez que segundo ele de acordo com a definição utilizada o descritor "Sensing Method Used" deveria ser designado de "Measurement Method Used". Outro caso particular foi o descritor "Frequency", que, segundo o investigador, deveria ser denominado "Sampling Frequency" uma vez que está definido como sendo o menor tempo possível entre uma observação e a próxima e segundo o investigador, este é uma frequência mínima, no entanto, um sensor geralmente pode ser configurado para recolher amostras em frequências mais baixas.

Por fim, foi ainda sugerido pelo investigador em alguns casos ser acrescentado à definição exemplos práticos, de forma a facilitar o entendimento do descritor.

No que respeita aos descritores escolhidos, selecionou-se um grupo de descritores mais geral, como é o caso do descritor "coordinates", "region", "distance", "working state", "start time" e "end time". Outro grupo de descritores está mais direcionado para os sensores e as suas características: "sensor", "device", "sensor data sheet", "platform", "system" e "unit". Os descritores "connected users" e "count" podem ser utilizados para descrever dados recolhidos através de aplicações móveis ou plataformas que exigem que os utilizadores estejam ligados a uma plataforma. E por fim, outro grupo de descritores, como são exemplos "measurement type", "sensing method used", "calibration type measurement" e "measurement range", usados para descrever o tipo de medição realizada.

Após a seleção e validação por parte do investigador dos descritores a utilizar na descrição dos dados, foi necessário formalizar os modelos de metadados M3-lite e SSN em ontologias no Dendro.

As ontologias desenvolvidas podem ser categorizadas como sendo ontologias ligeiras (do inglês *lightweight ontology)*, ontologias que apresentam formatos simplificados, compostas apenas por algumas classes e propriedades. A plataforma de gestão de dados (Dendro) é composta por ontologias ligeiras uma vez que são ontologias simples com restrições fracas o que facilita o processamento por máquinas e a incorporação das ontologias diretamente no Dendro (Castro, Silva, and Ribeiro 2014).

Para a formalização das ontologias M3-lite e SSN, no *Protégé* apenas se utilizaram *data properties*, uma vez que tal como os autores Castro, Silva, and Ribeiro (2014) referem que estas ontologias tornam-se alternativas viáveis para suportar o modelo de dados de um sistema de gestão de dados de investigação, uma vez que o Dendro, apresenta ainda algumas limitações e não suporta a utilização de classes no *Protégé*.

Outra perspetiva, esta mais relacionada com o curador de dados, é que a formalização de ontologias apenas com *data properties*, não "obriga" a que o curador de dados seja especialista na utilização do *software Protégé*, o que consequentemente contribui para que mais ontologias específicas em domínios sejam criadas. No caso por exemplo em que são os próprios investigadores a descrever os dados que produzem, na plataforma Dendro, o facto de serem utilizadas ontologias ligeiras e que não exigem que estes tenham um conhecimento especializado do *Protégé* contribuirá para que os investigadores não desistam de descrever os seus dados, uma vez que o processo é facilitado.

Para formalizar os descritores dos modelos de metadados M3-lite e SSN no Dendro, uma vez que ainda não estavam incorporados nas ontologias da plataforma Dendro foi necessário utilizar o *software Protégé*. Este *software* proporciona um conjunto de ferramentas que permite criar modelos de domínio, para além de ser um *software* gratuito e de código aberto (W3C

2009). Para o seu desenvolvimento usou-se como base as ontologias originais, por essa razão foi necessário importar as ontologias originais para o *Protégé* (ver Figura 17).

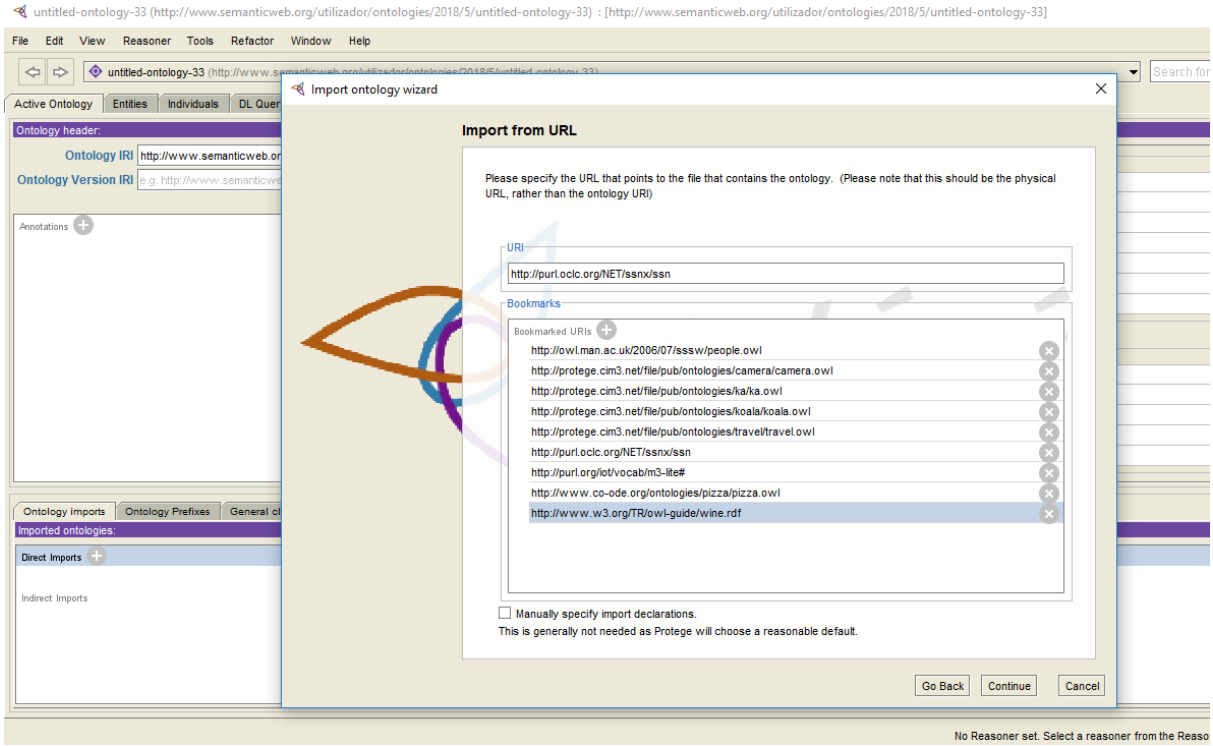


Figura 17 Importação da ontologia SSN para o Protégé.

Depois de importada a ontologia original no *Protégé* foram introduzidas as *data properties* como se pode ver na Figura 18, que são propriedades importantes que definem um valor sobre os dados. Sendo que foi necessário realizar o mesmo procedimento para os dois modelos de metadados.

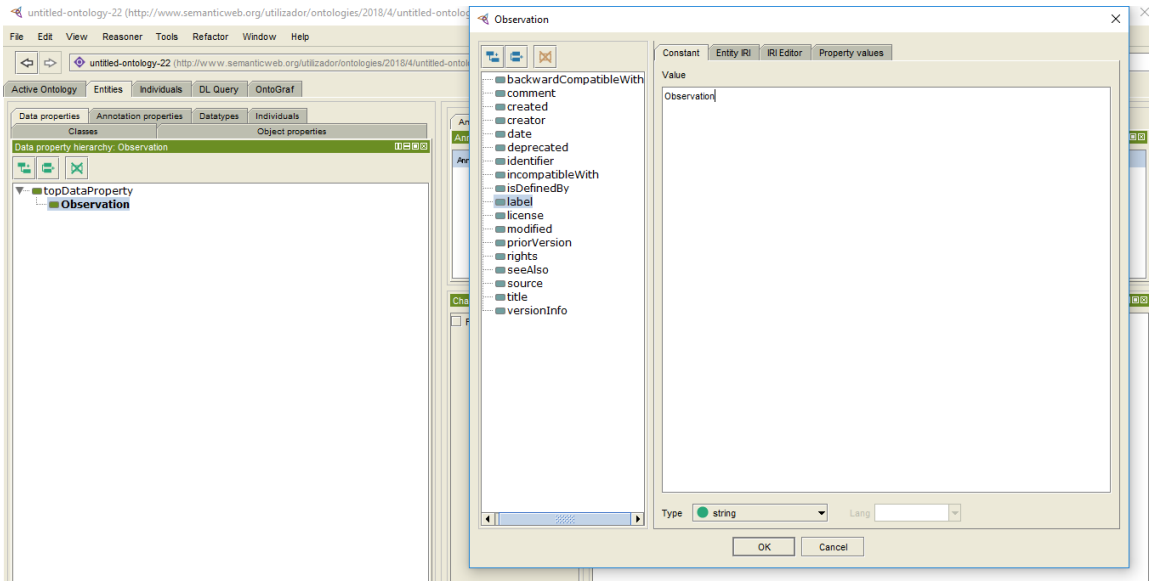


Figura 18 *Data Properties* da ontologia SSN.

Durante o processo de preparação da ontologia para o Dendro, foi necessário associar às *data properties* informação adicional que as pudesse caracterizar.

Assim, foram inseridas *labels* que permitem definir o formato como o descritor deve aparecer e *comments* em que se inseriu uma pequena descrição do descritor, informação essa retirada da ontologia original (ver Figura 19).

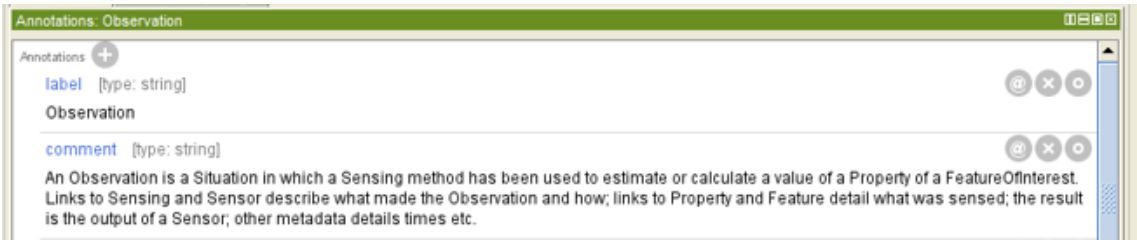


Figura 19 Atribuição de *labels* e *comments*.

Ainda relativamente às anotações utilizadas, foram também introduzidos dois *isDefinedBy*, um onde foi colocado o URL da ontologia original e outro que direciona para o conceito original da ontologia (ver Figura 20).

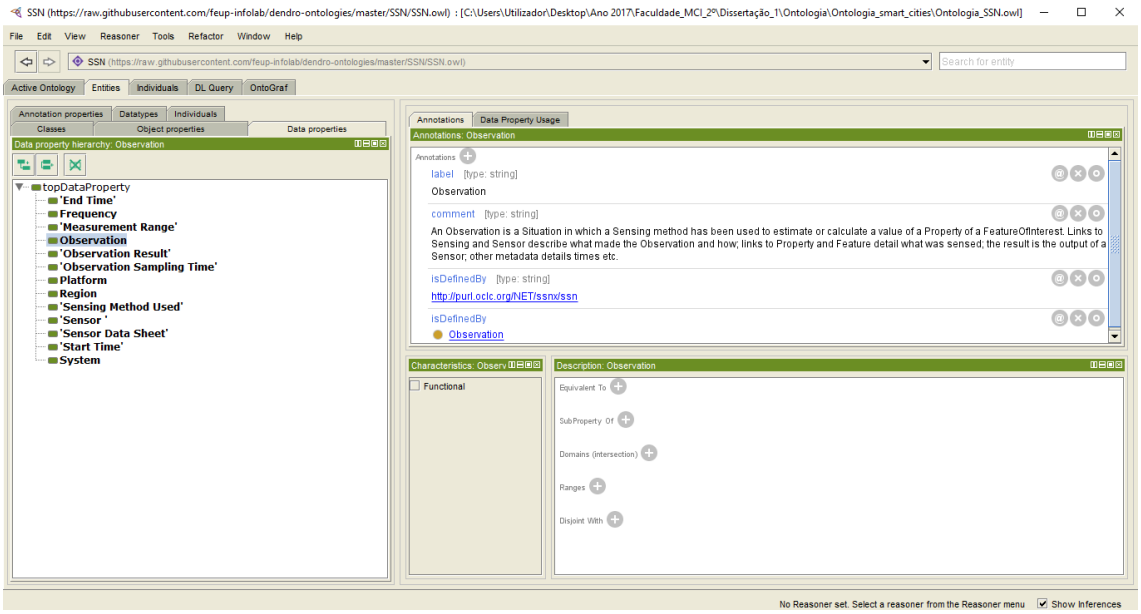


Figura 20 *Annotations* introduzidas na ontologia.

6.3 A Plataforma Dendro

O Dendro é uma plataforma web para armazenamento e descrição de dados de investigação. Esta é uma plataforma de código aberto, planeada para ajudar os utilizadores a armazenar e descrever os seus dados de forma colaborativa. O objetivo principal desta plataforma passa por auxiliar os investigadores a descreverem os seus dados facilitando a sua partilha e consequentemente permitir que os dados sejam reutilizados noutras investigações.

A plataforma Dendro foi construída desde logo com uma interface amigável para utilizadores sem qualquer conhecimento em gestão de dados. A plataforma permite criar ontologias de domínios específicos, com descritores criados sob medida e permite facilmente aos utilizadores combinar descritores de vários domínios (Silva et al. 2014).

6.4 Aplicação do modelo de metadados

Depois de terminada, a ontologia foi introduzida na plataforma de gestão de dados—Dendro. A plataforma já era composta por alguns modelos de metadados formalizados em ontologias como se pode ver na Figura 21. Alguns dos modelos de metadados são compostos por descritores mais genéricos, que podem ser aplicados a vários domínios, como por exemplo, Dublin Core, Friend of a Friend, entre outros. Outros modelos de metadados que compõem a plataforma de gestão de dados são formados por descritores aplicáveis a domínios específicos, como por exemplo, Biological Oceanography e Hydrogen Generation.

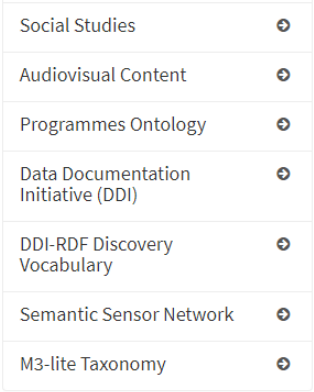


Figura 21 Exemplos de conjuntos de descritores incorporados no Dendro.

Na Figura 22 é possível ver a forma como os descritores são apresentados no Dendro. Os descritores estão ordenados alfabeticamente, e cada um tem associada uma descrição que auxilia os utilizadores a entender o que se pretende descrever utilizando aquele descritor. Este comentário foi definido anteriormente na *annotation property*, no *Protégé*.

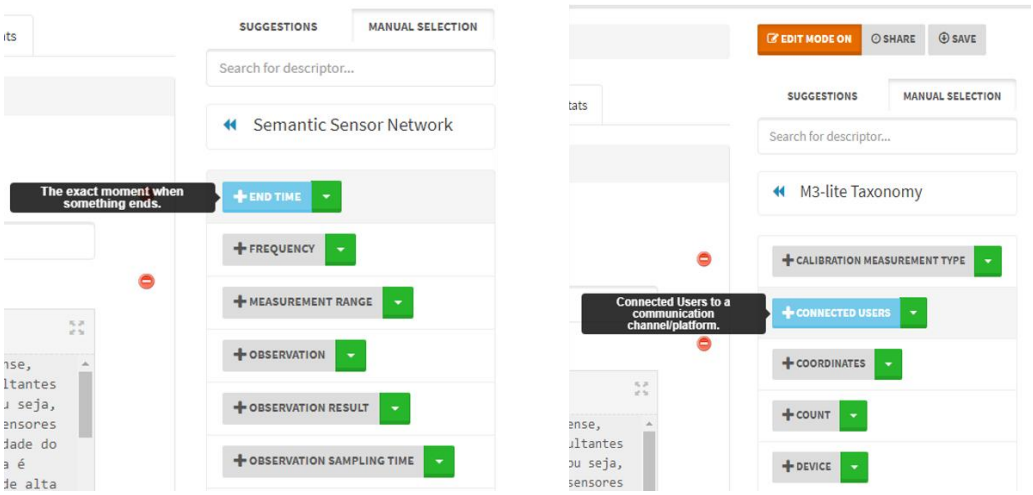


Figura 22 Exemplos de descritores e comentários associados.

Após serem inseridas as ontologias na plataforma Dendro, verificaram-se se os descritores estavam a funcionar corretamente.

Antes de passar à fase de descrição dos dados foi necessário analisar os descritores já existentes na plataforma Dendro e a partir dessa análise foram selecionados os que mais se adequavam ao conjunto de dados. Ainda nesta fase foram enviados os descritores ao investigador responsável pelo projeto, de forma a obter um *feedback* acerca dos descritores selecionados. O investigador recebeu um documento MSWord com todos os descritores selecionados e as respetivas definições para cada um, definições essa retiradas das ontologias originais, uma vez

que não foram criadas ontologias, foram apenas reutilizados alguns descritores de ontologias já existentes.

A maioria dos comentários realizados pelo investigador, foi ao nível das definições dos descritores. Percebeu-se que para o investigador, a definição de cada descritor não eram suficientes para o correto preenchimento dos descritores. Por outro lado, noutras situações segundo o investigador nem tudo o que compõem a definição é necessário para o contexto das *smart cities*.

Para além dos descritores selecionados, foi ainda proposto pelo investigador, o acréscimo de mais um descritor "Sensor Data Sheet", uma vez que segundo o investigador, outros investigadores que pretendiam utilizar os dados pediram as folhas de dados dos sensores.

Estas folhas de dados dos sensores são compostas por informações específicas dos sensores utilizados na recolha dos dados e tratam-se de documentos em formato *.pdf que contêm estas informações. Ou seja, para o preenchimento deste descritor seria necessário introduzir no campo de descrição os documentos *.pdf, no entanto a plataforma de gestão de dados não está preparada para a introdução de documentos *.pdf e ao fazer a publicação dos dados os documentos *.pdf poderiam ser perdidos. De forma a contornar esta situação, optou-se por introduzir no Dendro, dentro da pasta geral "Dados_USense", uma pasta, ao mesmo nível das pastas correspondentes às estações, denominada "Sensor Data Sheets" que continha as folhas de dados dos sensores. E no descritor "Sensor Data Sheets" colocou-se o nome dos documentos com indicação de que podem ser encontrados na pasta denominada "Sensor Data Sheets".

Relativamente ao descritor "Universe", o investigador não percebeu o contexto de utilização desse descritor, no entanto optou-se por manter o descritor uma vez que este permite contextualizar o universo a ser analisado, que neste caso corresponde a dados recolhidos por 13 sensores, que se encontram dispostos em 23 locais da cidade do Porto.

No que se refere aos restantes descritores, o investigador não demonstrou nenhum entrave na utilização dos mesmos, pelo que se assumiu a total utilização dos descritores previamente selecionados.

Após tudo estar validado passou-se à fase seguinte que consistia na ingestão dos dados UrbanSense na plataforma e a respetiva descrição.

Para dar início à descrição dos dados UrbanSense, foi criado um projeto denominado "UrbanSense" onde foram inseridas algumas informações gerais acerca do projeto, nomeadamente uma pequena descrição acerca do projeto. De seguida foi realizado o *upload* dos dados a serem descritores, tratando-se de uma grande quantidade de dados o carregamento foi demorado. Ao longo desta fase, o Dendro foi apresentando alguns desafios, uma vez que os

dados introduzidos tinham um tamanho extenso e habitualmente a plataforma lida com tamanhos menores.

Dado que numa fase anterior já se tinham organizado os dados por pastas, essa organização manteve-se ao inserir o *dataset* no Dendro, tal como se pode ver na Figura 23.

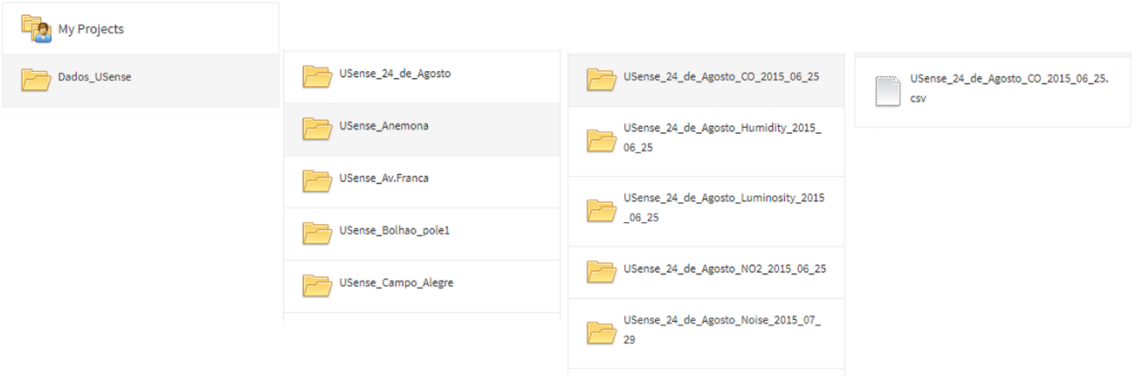


Figura 23 Esquema representativo da organização das pastas no Dendro.

Uma vez introduzidos os dados na plataforma, iniciou-se o processo de descrição. Para descrever os dados, como se pode ver na Figura 24, selecionaram-se os respetivos descritores que se encontram listados do lado direito da imagem. Na parte central da imagem que corresponde à área de descrição, é possível ver alguns dos descritores já selecionados e preenchidos.

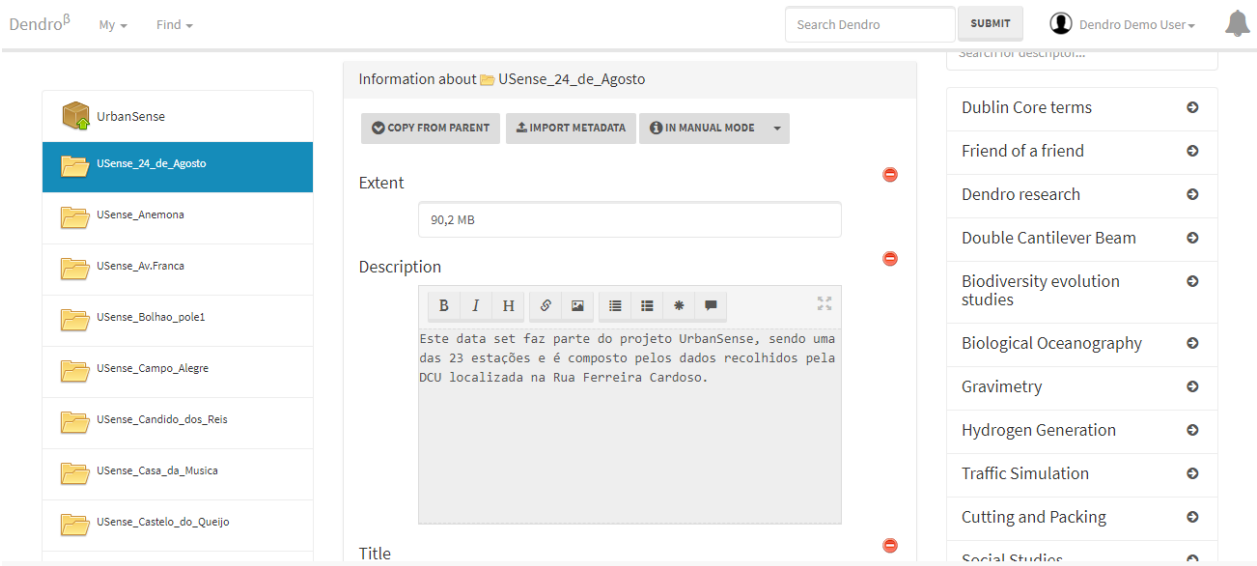


Figura 24 Exemplos de descritores selecionados e preenchidos para a descrição do *data set*.

É importante referir que para a descrição dos dados UrbanSense foram utilizados descritores de outros modelos de metadados, pois muitos dos descritores adequados para a descrição dos *data sets* já existiam na plataforma Dendro.

A descrição dos *data sets* foi feita ao nível da pasta principal denominada "Dados USense", onde foram utilizados 21 descritores. Num segundo nível optou-se por fazer uma descrição ao nível das pastas referentes às estações onde estão localizadas as DCUs, totalizando 23 pastas e 18 descritores para cada pasta. Por uma questão de tempo, não se optou por fazer a descrição ao nível dos ficheiros, uma vez que se tratavam de 328 ficheiros e no Dendro existe uma relação hierárquica entre os ficheiros. Assim, os ficheiros que estão dependentes de um nível acima recebem a descrição desse nível, desta forma as pastas que estão abaixo das pastas referentes às estações (que foram descritas) recebem essa descrição. Para descrever os dados UrbanSense selecionaram- se os seguintes descritores dentro dos vários modelos de metadados (ver Tabela 5):

Tabela 5 Descritores selecionados para descrever os dados UrbanSense

Modelo de metadados	**Descritores**
Dublin Core	Access Rights
	Contributor
	Creator
	Description
	Extent
	Format
	Identifier
	Is Part Of
	Language
	License
	Publisher
	Relation
	Spatial Coverage
	Subject
	Title
	Type
Semantic Sensor Network (SSN)	Measurement Range
	Region
	Sensor
	Unit
M3-lite	Coordinates

	Working State
Data Documentation Initiative (DDI)	Methodology
	Universe
DDI- RDF Discovery Vocabulary	Start Date
	End Date
Friend of a Friend	Personal Mailbox

A maioria dos descritores selecionados são descritores Dublin Core, pois este é um modelo de metadados que pode ser facilmente adaptado a vários domínios e permite descrever vários tipos de recursos e formatos de documentos. A utilização dos descritores Dublin Core demonstrou-se eficaz para descrever informação mais geral do projeto.

Relativamente a questões mais técnicas foi necessária a seleção de descritores do SSN, M3-lite e DDI uma vez que se conclui ser necessário incluir na descrição informação ao nível das unidades de medida, dos sensores utilizados na recolha e período de tempo de recolha de dados correspondente a cada DCU.

6.5 Publicação dos dados no repositório do INESC TEC

Uma correta organização dos dados e consequente descrição dos mesmos, garantindo sempre a sua preservação com o objetivo à partilha são tarefas essenciais para garantir a possibilidade de reutilização dos dados.

A partilha dos dados UrbanSense foi realizada no repositório do INESC TEC, um repositório de dados que agrega diversos conjuntos de dados produzidos pelos investigadores do INESC TEC e dos seus parceiros.

Para a tarefa dedicada à publicação dos dados, foi necessário ter em conta vários aspetos, uma vez que existiam duas alternativas para a publicação dos *data sets*. Uma das alternativas permitia que a publicação dos *data sets* fosse realizada diretamente da plataforma Dendro para o repositório. Outra das alternativas implicava que a publicação fosse realizada à parte do Dendro e exigia o preenchimento dos metadados novamente no repositório.

Cada uma das alternativas tinham algumas implicações diretas na exposição dos *data sets*, uma vez que se trata de um conjunto de subpastas. A primeira alternativa implicava a perda dos metadados referentes aos *data sets* das estações, ou seja, o investigador quando acedesse às pastas referentes às estações, não teria nenhuma informação que contextualizasse aquele *data set* e implicava ainda que estivesse disponível um único *data set*, a pasta principal e dentro dessa todas as restantes pastas. Consequentemente esta alternativa implicava que no futuro um

investigador que quisesse utilizar por exemplo os dados recolhidos na Rua das Flores, fosse "obrigado" a descarregar o *data set* completo.

A segunda alternativa, implicava o preenchimento dos metadados novamente no repositório, quanto à disposição dos *data sets* esta permitia que fossem introduzidos os 23 *data sets* referentes às estações e que para cada um deles fosse introduzida uma pequena descrição em que poderiam ser disponibilizadas informações específicas daquele *data set*.

Desta forma a solução mais viável passou pela publicação dos *data sets* diretamente no repositório do INESC TEC. Para a realização da publicação, foi necessário extrair do Dendro um ficheiro *.zip com todas as pastas que estavam na plataforma e ainda foram gerados dois outros ficheiros, um ficheiro *.txt e um *.rdf ambos com os metadados descritos no Dendro, como se pode ver na Figura 25.

```
dcterms:http://purl.org/dc/terms/
foaf:http://xmlns.com/foaf/0.1/
ddr:http://dendro.fe.up.pt/ontology/0.1/
rdf:http://www.w3.org/1999/02/22-rdf-syntax-ns#
nie:http://www.semanticdesktop.org/ontologies/2007/01/19/nie#
disco:http://rdf-vocabulary.ddialliance.org/discovery#
ddiup:http://dendro.fe.up.pt/ontology/ddiup#
m3lite:https://raw.githubusercontent.com/feup-infolab/dendro-ontologies/master/M3-LITE/M3-lite.owl#
ssn:https://raw.githubusercontent.com/feup-infolab/dendro-ontologies/master/SSN/SSN.owl#

./Dados_USense
| +Extent(dcterms:extent) : 3,06 GB
| +Access Rights(dcterms:accessRights) : Acesso aberto
| +License(dcterms:license) : Creative Commons Atribution Share-Alike
| +Description(dcterms:description) : Estes dados são resultados do projeto UrbanSense, pertencem ao domínio ambiental. Os dados resultantes das investigações são dados observacionais,
ou seja, dados recolhidos em tempo real pelos vários sensores que se encontram dispostos em 23 locais da cidade do Porto. A calibração dos sensores da plataforma é realizada utilizando
um conjunto de sensores de alta qualidade para criar modelos de calibração baseados em regressões lineares e não-lineares.
Os dados analisados neste projeto, não sofrem qualquer tipo de processamento, tratando-se de dados brutos.
Os sensores que compõem as unidades de recolha de dados (DCU) da Plataforma UrbanSense são: Weather station; Noise; CO; Temperature; Precipitation; Luminosity; Wind direction; Humidity;
Solar radiation; Particles 2; Particles 1; O3; NO2 and Wind speed.
As estações onde se encontram os vários sensores estão localizadas na Rua das Flores; Damião de Góis; Reitoria; Praça do Marques- pole1; Combatentes; Estádio do Dragão; Bolhão- pole1;
Avenida de França; Casa da Música; Trindade; Castelo do Queijo; Anemona; Campo Alegre; D.Manuel II; Praça da Galiza; Hospital de S.João; Congregados; Cândido dos Reis; Praça da Liberdade
(Cardosas); 24 de Agosto (Ferreira Cardoso); Fundação Serralves; FEUP (P1) e Praça Velasquez.
| +Publisher(dcterms:publisher) : Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência - INESC TEC
| +Title(dcterms:title) : Dados UrbanSense
| +Contributor(dcterms:contributor) : Comissão Europeia
| +Creator(dcterms:creator) : Ana Cristina Costa Aguiar
| +Identifier(dcterms:identifier) : https://futurecities.up.pt/site/hybrid-sensor-networking-testbed/
| +Language(dcterms:language) : Português
| +Relation(dcterms:relation) : Artigos relacionados com a plataforma UrbanSense:
```

Figura 25 Excerto do ficheiro *.txt com os metadados preenchidos no Dendro.

Uma segunda etapa da publicação implicou o preenchimento de um documento disponibilizado pela responsável pelas publicações no repositório. Nesse documento disponível através do link https://docs.google.com/spreadsheets/d/1hpl_BRKno-JekMoyk9ceyhu_6bRN_NCE5XCtL1o96wA/edit?usp=sharing foram preenchidos os metadados gerais referentes ao projeto, a maioria deles eram descritores do modelo de metadados Dublin Core e foi possível ainda acrescentar descritores de outros modelos de metadados anteriormente utilizados, como o caso do M3-lite, SSN e DDI (ver Figura 26).

Dataset | Groups | Activity Stream

UrbanSense environmental monitoring

The UrbanSense project is the environmental monitoring part of the Smart City initiative at the city of Porto, Portugal. This dataset contains observational data collected at 23 locations: Rua das Flores; Damião de Góis; Reitoria; Praça do Marquês-pole1; Combatentes; Estádio do Dragão; Bolhão-pole1; Avenida de França; Casa da Música; Trindade; Castelo do Queijo; Anemona; Campo Alegre; D.Manuel II; Praça da Galiza; Hospital de S.João; Congregados; Cândido dos Reis; Praça da Liberdade (Cardosas); 24 de Agosto (Ferreira Cardoso); Fundação Serralves; FEUP (P1); Praça Velasquez. The data collection unit in each location is composed by sensors of several kinds: Weather station; Noise; CO; Temperature; Precipitation; Luminosity; Wind direction; Humidity; Solar radiation; Particles 2; Particles 1; O3; NO2 and Wind speed. The sensors' calibration is done with a set of high quality sensors to create calibration models based on linear and nonlinear regressions. Data in this project is not subject to any kind of processing, everything is raw data.

Field	Value
Author	Ana Cristina Costa Aguiar
Last Updated	23 de Julho de 2018, 10:28 (UTC+01:00)
Created	21 de Junho de 2018, 17:03 (UTC+01:00)
dc.Contributor	European Commission
dc.Coverage.Spatial	Porto, Portugal
dc.Coverage.Temporal	22/06/2014 a 26/08/2017
dc.Format	*.csv; *.pdf
dc.Format.Extent	3.06GB
dc.Identifier	https://futurecities.up.pt/site/hybrid-sensor-networking-testbed/
dc.Language	PT

dc.Relation	1. Yunior Luis, Pedro M. Santos, Tiago Lourenco, Carlos Pérez-Penichet, Tânia Calcada, and Ana Aguiar. 2012. UrbanSense: an Urban-scale Sensing Platform for the Internet of Things. 2016 IEEE Smart Cities Conference (ISC2). DOI: 10.1109/ISC2.2016.7580869 2. Pedro M. Santos, Tânia Calçada, Diogo Guimarães, Tiago Condeixa, Susana Sargento, Ana Aguiar, João Barros. 2015. Demo: Platform for Collecting Data From Urban Sensors Using Vehicular Networking. DOI: 10.1145/2789168.2789169
dc.Type	Observational data
ddi.InstrumentType	Static sensors
ddi.Methodology	The UrbanSense Platform is made up of Data Collection Units (DCU) in which complex communication infrastructures exist in the urban area, together with a Delay Tolerant Network (DTN), in which the sensors of the infrastructure PortoLivingLab through the wireless connection points reach the knots backhaul. In the case of the UrbanSense project, the collected data is transferred to the cloud backend server through Wi-Fi APs.

ddi.Universe	The project comprises data collected by 13 sensors, which are arranged in 23 locations in the city of Porto.
m3-lite.Working State	Disabled
qu.Unit	CO - volts; Humidity - %; Luminosity - volts; NO2 - volts; Noise - dB; O3 - volts; Particles 1 - counts; Particles 2 - counts; Solar radiation - volts; Temperature - C; Wind direction - degrees; Wind speed - rps (1rps=1,492miles/h); Precipitation - number of buckets (1 bucket=0,2794mm)
ssn.SensorDataSheet	Documents with information about the sensors available in the folder called "Sensor Data Sheet"

Figura 26 Metadados preenchidos no repositório do INESC TEC.

Tal como referido anteriormente, como se optou pela publicação dos dados diretamente no repositório do INESC TEC, foram depositados no total 23 *data sets* referentes às estações, mais um *data set* denominado "Sensor Data Sheet" composto por documentos *.pdf com informações específicas sobre os sensores e mais dois ficheiros com todos os metadados descritos anteriormente no Dendro.

A vista geral do depósito dos *data sets* no repositório é apresentada na Figura 27.

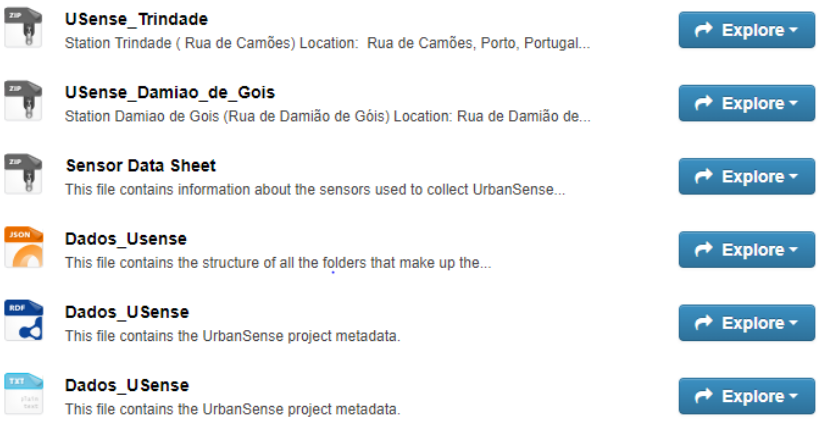


Figura 27 Excerto da vista geral do depósito dos *data sets* UrbanSense.

Se um utilizador selecionar um *data set* específico por exemplo "USense_Trindade" pode visualizar uma pequena descrição referente ao *data set* (ver Figura 28), para aceder aos metadados completos é necessário aceder ao ficheiro *.txt ou *.rdf.

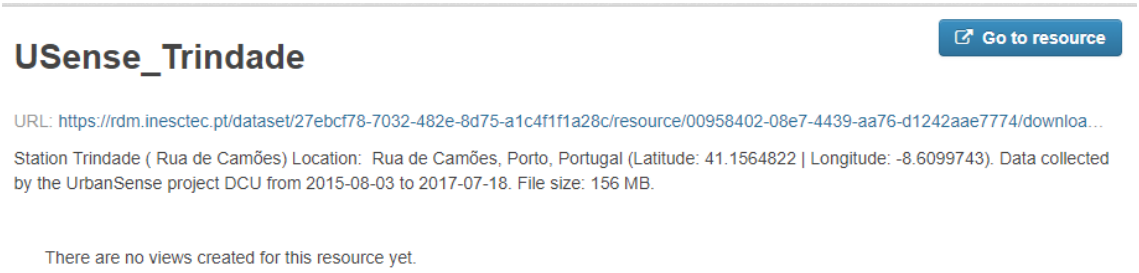


Figura 28 Descrição disponível no *data set*.

7 Avaliação dos resultados

O foco deste trabalho centrou-se na gestão de dados de investigação do projeto UrbanSense, uma vez que estes dados têm um grande potencial de reutilização e por isso é importante investir na sua descrição. No entanto, é também importante avaliar a sua eficácia num contexto prático. Ou seja, testar a descrição realizada e consequentemente a sua aplicabilidade. O objetivo centrou-se em perceber junto de um investigador pertencente ao domínio de investigação em causa e um investigador que não é do domínio em questão, mas que, na sua prática profissional, executa tarefas de gestão e curadoria de dados de investigação, se através do repositório onde foram divulgados os dados, num contexto de reutilização, os descritores utilizados seriam suficientes para os investigadores se contextualizarem e consequentemente reutilizarem os dados.

Para contribuir para esta pequena avaliação optou-se por convidar um investigador que não possuía qualquer tipo de conhecimento acerca do projeto UrbanSense, para resultar numa avaliação mais concreta, uma vez que o foco da partilha destes dados é para outros investigadores que não participaram no projeto e convidou-se ainda um investigador que não pertence ao domínio, mas que está familiarizado com práticas de curadoria de dados, para que fosse possível comparar duas avaliações.

A experiência teve início com uma explicação do contexto desta dissertação e da necessidade de obter uma avaliação por parte do investigador. Na consequência de se optar por selecionar investigadores que não pertenciam ao projeto, nada se explicou acerca do mesmo, de forma a não enviesar os comentários dos investigadores.

Na análise do investigador do domínio, logo desde início referiu que a informação que estava na página principal estava bem sintetizada e era suficiente para perceber o que eram os dados, onde estavam e o que recolhiam os sensores. O mesmo aconteceu para o investigador que não pertencia ao domínio ambiental, que referiu que os descritores selecionados eram apropriados face ao contexto dos dados do projeto.

O investigador do domínio, referiu ainda que a informação disponibilizada era mais que suficiente para a contextualização e que existia mesmo informação que nem seria necessária. Uma questão apontada pelo investigador foi referente às datas em que os dados começaram a ser recolhidos, que segundo ele apesar de estar disponível em cada *data set* as datas em que os dados daquele *data set* foram recolhidos, o mesmo sentiu necessidade de ter as datas na descrição principal. No entanto, as datas estavam disponíveis no descritor "dc.Coverage.Temporal: 22/06/2014 a 26/08/2017" o que pode levar a concluir que o descritor selecionado não seja o mais apropriado. Ainda foi apontada outra questão por parte do investigador, que referiu fazer-lhe alguma confusão a unidade de medida "count" nas partículas 1 e 2 (que entende como sendo partículas primárias e secundárias), o que lhe levou a questionar

de que tamanho seriam, apesar desta ser uma questão que ultrapassa o curador de dados, uma vez que não é do domínio e, portanto, não sabe informação tão específica, deve ser uma questão a ter em conta junto dos investigadores do projeto. Por fim, o investigador imaginando que queria utilizar estes dados num outro projeto, questionou se poderia utilizar os dados na sua investigação e poderiam ser publicados, pelo que lhe foi respondido que os dados estão em acesso aberto, no entanto para questões mais específicas está disponibilizado o contacto do responsável pelo projeto.

Relativamente ao outro investigador que se demonstrou mais familiarizado com os termos metadados e descritores, referiu que dada a dimensão do projeto certamente muita mais informação haveria a dar, no entanto afirmou que esse seria o desafio do curador que demonstrou ter sentido de síntese na seleção dos descritores, uma vez que não conhecia profundamente o projeto.

Em suma, a análise do investigador foi ao encontro do pretendido com esta avaliação uma vez que o mesmo referiu que a descrição realizada poderia ser útil por dois lados, para os potenciais utilizadores do projeto, como o caso dos investigadores que realizaram esta avaliação, e ainda para os próprios investigadores que poderão eventualmente esquecer-se de pormenores do projeto e esta descrição funcionará como anotação capaz de dar contexto.

Uma questão importante que ambos os investigadores referiram é que a descrição realizada não era demasiado redutora, que segundo eles poderia condicionar a interpretação, nem demasiado extensa, pormenor que foi tido em conta aquando da seleção dos descritores.

8 Conclusões e reflexão crítica

8.1 Resumo dos resultados obtidos

O acesso aos dados de investigação é ainda hoje um tema inquietante para muitos investigadores e entidades envolvidas nos projetos de investigação, quer por não estarem conscientes dos aspetos ligados à preservação, quer ao acesso aos dados. É por isso importante, consciencializar as instituições envolvidas nos projetos de investigação das vantagens resultantes do acesso e partilha dos dados de investigação. Grande parte dos dados produzidos em investigações de pequenos grupos têm grande valor e potencial de reutilização e quando disponibilizados podem resultar em novas investigações. Por isso, em vez de, como acontece muitas vezes os dados resultantes de investigações serem utilizados apenas por investigadores que trabalharam no projeto, devem ser partilhados com terceiros para que desta forma sejam evitadas duplicações e os próprios projetos ganhem valor.

Ao longo deste trabalho ficou visível a progressão na preocupação quanto à gestão de dados, apesar do projeto UrbanSense não possuir um Plano de Gestão de Dados e estratégias de partilha dos dados, os investigadores demonstraram consciência da potencialidade de reutilização dos dados que foram produzidos no projeto.

Neste sentido, o trabalho desenvolvido nesta dissertação, pretendeu apoiar a gestão de dados de investigação, focando-se na análise, organização, descrição e publicação dos dados recolhidos no projeto UrbanSense, num repositório de dados de investigação de acesso aberto. Desde logo, se iniciou o processo desenvolvendo um Plano de Gestão de Dados. Uma vez que o mesmo não continha um PGD sentiu-se a necessidade de criar um plano que orientasse a gestão dos dados.

Depois de aceder aos dados, verificou-se que era uma grande quantidade de dados e foi nesta fase que começaram a surgir algumas dificuldades, já referenciadas ao longo do trabalho, desde a obtenção dos dados através da plataforma UrbanSense, que devido à sua dimensão nem sempre a plataforma devolvia os dados pretendidos, até à descrição dos dados. Sendo importante referir que se optou por publicar os dados sem qualquer avaliação da qualidade dos mesmos, uma vez que em conjunto com os investigadores se percebeu que a avaliação deveria ser feita pelo investigador que pretende utilizar os dados tendo em conta a finalidade da sua utilização.

Para auxiliar a questão da descrição dos dados, que é essencial para que estes sejam compreensíveis e sejam partilhados e reutilizados por outros investigadores, foram selecionados descritores de vários modelos de metadados com o intuito de descrever os dados UrbanSense utilizando a plataforma Dendro.

Nesta fase, foram ainda reutilizados descritores de dois modelos de metadados que não integravam a plataforma Dendro e que seriam úteis para a descrição quer do projeto UrbanSense, quer numa perspetiva futura de dados relacionados com as *smart cities*.

O envolvimento dos investigadores é fundamental para se perceber o domínio e todo o processo de recolha dos dados, pelo que tanto as entrevistas realizadas aos investigadores, quanto a avaliação por parte dos investigadores da seleção dos descritores que iriam ser utilizados na descrição dos dados se tornaram tarefas importantes para a gestão dos dados. Nem sempre há envolvimento dos investigadores com o curador de dados, no entanto a realização deste estudo veio provar que esta relação deve ser vinculada, tanto para a correta descrição dos dados, uma vez que o curador não está dentro do domínio, quanto para o sucesso dos dados e da sua reutilização.

A avaliação final realizada por um investigador do domínio e um investigador que não era do domínio veio comprovar que os dados aqui tratados, têm potencial, e podem ser utilizados em diversos âmbitos como prevenções, perceber tendências, quer ao nível do clima, meio ambiente, possíveis anomalias que exigem intervenções urbanas, planeamento urbano, entre outros. Esta avaliação dos investigadores também fundamentou a descrição realizada, uma vez que os investigadores através dos metadados conseguiram contextualizar-se no projeto UrbanSense.

Em suma, espera-se que esta dissertação contribua para a consciencialização dos investigadores e entidades envolvidas da necessidade de gerir os dados por eles produzidos e do seu potencial.

8.2 Trabalho futuro

Na perspetiva do trabalho realizado e da sua importância nos projetos de investigação e tendo em conta a abrangência do projeto, uma vez que juntamente com a plataforma UrbanSense, fazem parte do projeto PortoLivingLab, outras duas plataformas a BusNet e o SenseMyCity, são claras as possibilidades futuras relativamente à gestão destes dados.

Uma vez que os investigadores demonstraram desde logo interesse na gestão dos dados de investigação das restantes plataformas, propôs-se ao longo deste trabalho incorporar na plataforma Dendro dois modelos de metadados (M3-lite e SSN). O objetivo centrou-se na reutilização de metadados já existentes nestes modelos, mas que não estavam incorporados na plataforma de gestão de dados, Dendro

Para formalizar estes modelos de metadados no Dendro, foram analisados e selecionados descritores desenvolvidos para o domínio da IoT e para a descrição de sensores

que poderiam vir a ser utilizados na descrição dos dados quer do projeto BusNet, quer do projeto SenseMyCity.

Desta forma, futuramente aquando da descrição dos dados das restantes plataformas o curador de dados terá à sua disposição dois modelos de metadados que poderão ser utilizados em conjunto com outros modelos na descrição dos dados.

Espera-se com a realização desta dissertação e com a descrição e publicação dos dados que os investigadores vejam resultados positivos nos seus projetos, que os mesmos tenham mais visibilidade e que haja uma maior divulgação dos dados. Consequentemente, que haja uma maior sensibilização para a cultura de gestão de dados, uma vez que o processo de gestão de dados exige muito esforço e não pertence só aos curadores de dados este trabalho, pois tem que ser um trabalho de envolvência entre o curador e os investigadores.

Referências Bibliográficas

Bicarregui, Juan, Norman Gray, Rob Henderson, Roger Jones, Simon Lambert, and Brian Matthews. 2013. "Data Management and Preservation Planning for Big Science." *International Journal of Digital Curation* 8 (1):29–41. https://doi.org/10.2218/ijdc.v8i1.247.

Castro, João Aguiar, João Rocha da Silva, and Cristina Ribeiro. 2014. "Creating Lightweight Ontologies for Dataset Description: Practical Applications in a Cross-Domain Research Data Management Workflow," 0–3.

Costa, Maira Murrieta, and Murilo Bastos da Cunha. 2014. "O Bibliotecário No Tratamento de Dados Oriundos Da E-Science: Considerações Iniciais." *Perspectivas Em Ciência Da Informação* 19:189–206. https://doi.org/10.1590/1981-5344/1900.

Curtin University. 2018. "Research Data Management: Data Description." 2018. http://libguides.library.curtin.edu.au/c.php?g=202401&p=1333152.

Digital Curation Centre. 2006. "Curating E-Science Data." *DCC Briefing Papers*, no. October. http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/curating-e-science-data.

———. 2018a. "Data Management Plans." 2018. http://www.dcc.ac.uk/resources/data-management-plans.

———. 2018b. "DCC Curation Lifecycle Model." 2018. http://www.dcc.ac.uk/resources/curation-lifecycle-model.

———. 2018c. "What Is Digital Curation?" 2018. http://www.dcc.ac.uk/digital-curation/what-digital-curation.

European Comission. 2016. "H2020 Programme Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020." http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

Eynden, Veele Van den, Louise Corti, Martthew Woollard, Libby Bishop, and Laurence Horton. 2011. *Managing and Sharing Data - Best Practice For Researchers. UK Data Archive University of Essex*. 3rded. UK Data Archive; University of Essex Wivenhoe; Park Colchester; Essex; CO4 3SQ. https://doi.org/10.1007/978-1-84996-226-1.

Furtado, Filipe, Pedro Príncipe, and José Carvalho. 2017. "Kit Sobre Dados de Investigação RCAAP." http://repositorium.sdum.uminho.pt/handle/1822/46351.

Gattelli, Rúbia Tatiana. 2015. "Gestão de Dados de Investigação no domínio da Oceanografia Biológica: Criação E Avaliação de Um Perfil de Aplicação Baseado Em Ontologia." Universidade do Porto. https://doi.org/10.1002/mrd.22467.

Harvard Library. 2018. "Data Management Plans." 2018. https://guides.library.harvard.edu/dmp.

Hey, Tony, and Anne Trefethen. 2003. "E-Science and Its Implications." *The Royal Society*, no.

June:1809–25. http://rsta.royalsocietypublishing.org/content/roypta/361/1809/1809.full.pdf.

Higgins, Sarah. 2018. "What Are Metadata Standards." 2007. 2018. http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/what-are-metadata-standards#top.

Horton, Laurence, Veerle V A N Den, Louise Corti, and Libby Bishop. 2011. "Data Management Recommendations for Research Centres and Programmes," no. March.

Joint Information Systems Committee. 2014. "How Jisc Is Helping Researchers." 2014. http://webarchive.nationalarchives.gov.uk/20140702225057/http://www.jisc.ac.uk/whatweedo/campaigns/res3/jischelp.aspx#ideas.

Luis, Yunior, Pedro M. Santos, Tiago Lourenço, Carlos Pérez-Penichet, Tânia Calçada, and Ana Aguiar. 2016. "UrbanSense: An Urban-Scale Sensing Platform for the Internet of Things." *IEEE 2nd International Smart Cities Conference: Improving the Citizens Quality of Life, ISC2 2016 - Proceedings*. https://doi.org/10.1109/ISC2.2016.7580869.

National Science Board. 2005. *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*. *National Science Board*. https://doi.org/citeulike-article-id:7781668.

OECD. 2007. "OECD Principles and Guidelines for Access to Research Data from Public Funding." *Data Science Journal*. Paris. https://doi.org/10.2481/dsj.6.OD4.

Pennock, Maureen. 2007. "Digital Curation: A Life-Cycle Approach to Managing and Preserving Usable Digital Information." *Library and Archives Journal* 1 (January):1–3. http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf.

Ribeiro, Cristina, Ricardo Carvalho Amorim, João Rocha da Silva, João Aguiar Castro, and João Correia Lopes. 2016. "Projeto TAIL — Gestão de Dados de Investigação Da Produção Ao Depósito E À Partilha (Resultados Preliminares)." *Cadernos BAD*, no. Julho:256–64. http://www.bad.pt/publicacoes/index.php/cadernos/article/viewFile/1603/pdf.

Rodrigues, Eloy, Ricardo Saraiva, Cristina Ribeiro, and Eugénia Matos Fernandes. 2010. "Os Repositórios de Dados Científicos: Estado Da Arte." http://repositorium.sdum.uminho.pt/handle/1822/10830.

Santos, Pedro M, G P Rodrigues, Susana B Cruz, Tiago Lourenço, Pedro M Orey, Yunior Luis, Susana Sargento, and Ana Aguiar. n.d. "PortoLivingLab : An IoT-Based Sensing Platform for Smart Cities." *IEEE INTERNET OF THINGS JOURNAL*, 1–9.

Silva, João Rocha da, João Aguiar Castro, Cristina Ribeiro, and João Correia Lopes. 2014. "The Dendro Research Data Management Platform: Applying Ontologies to Long-Term Preservation in a Collaborative Environment."

Simms, Stephanie, and Sarah Jones. 2017. "Next-Generation Data Management Plans: Global, Machine-Actionable, FAIR." *International Journal of Digital Curation* 12 (1):36. https://doi.org/10.2218/ijdc.v12i1.513.

University of Minnesota. 2018. "Research Data Management (RDM)." 2018. http://libguides.d.umn.edu/c.php?g=307731&p=2409893.

W3C. 2009. "Protégé." 2009. https://www.w3.org/2001/sw/wiki/Protege.

Anexo I. Plano de Gestão de Dados – UrbanSense

O Plano de Gestão de Dados foi criado usando o DMPonline

Criador: patriciagsd03@gmail.com

Afiliação: University of Porto (Universidade do Porto)

Financiador: Comissão Europeia

Resumo do Projeto:

A plataforma UrbanSense faz parte do projeto Future Cities, que objetivou a transformação da cidade do Porto numa Smart City, intitulado de PortoLivingLab.
O PortoLivingLab é uma estrutura desenvolvida no âmbito das Smart Citys que apontam para a melhoria da qualidade de vida dos cidadãos, permitindo uma melhor compreensão dos processos urbanos da cidade do Porto e é composta por três plataformas de monitorização, o SenseMyCity, o BusNet e o UrbanSense.
A Plataforma UrbanSense é uma infraestrutura desenvolvida para monitorizar vários fenómenos ambientais, composta por Unidades de Recolha de Dados (Data Collection Unit - DCU) compostas por um conjunto de sensores que recolhem dados ambientais e que estão colocados em locais estratégicos na cidade do Porto.

Última modificação: 23-05-2018

Informação de Copyright:

UrbanSense

RECOLHA DE DADOS

Que dados irá recolher ou criar?

A plataforma UrbanSense, faz parte do projeto Future Cities, que objetivou a transformação da cidade do Porto numa Smart City, intitulada de PortoLivingLab.
O PortoLivingLab é uma plataforma desenvolvida para as Smart Citys que apontam para a melhoria da qualidade de vida dos cidadãos, permitindo uma melhor compreensão dos processos urbanos da cidade do Porto e é composta por três plataformas de monitorização, o SenseMyCity, o BusNet e o UrbanSense.
A Plataforma UrbanSense é uma infraestrutura que atua desde 2014, desenvolvida para monitorizar vários fenómenos ambientais, composta por 13 sensores, denominados Unidades de Recolha de Dados (DCU), que estão espalhados por 23 locais estratégicos na cidade do Porto, denominadas de estações.
As unidades que compõem a Plataforma UrbanSense são: Weather station; Noise; CO; Temperature; Precipitation; Luminosity; Wind direction; Humidity; Solar radiation; Particles 2; Particles 1; O3; NO2 and Wind speed.

As estações onde se encontram os vários sensores estão localizadas na Rua das Flores; Damião de Góis; Reitoria; Praça do Marques- pole1; Combatentes; Estádio do Dragão; Bolhão- pole1; Avenida de França; Casa da Música; Trindade; Castelo do Queijo; Anemona; Campo Alegre; D.Manuel II; Praça da Galiza; Hospital de S.João; Congregados; Cândido dos Reis; Praça da Liberdade (Cardosas); 24 de Agosto (Ferreira Cardoso); Fundação Serralves; FEUP (P1) e Praça Velasquez.

Como os dados serão recolhidos ou criados?

A Plataforma UrbanSense é composta por unidades estáticas de recolha de dados (DCU) em que são aproveitadas infraestruturas de comunicação complexas existentes em áreas urbanas, juntamente com uma Delay Tolerant Network (DTN), em que os sensores da infraestrutura PortoLivingLab através dos pontos de ligação sem fios alcançam os nós backhaul. No caso do projeto UrbanSense, os dados recolhidos são transferidos para o servidor backend cloud através APs Wi-Fi.
Depois de recolhidos os dados serão armazenados em bases de dados e posteriormente passarão por uma análise e publicação num repositório. Para a realização dessa análise é necessário aceder aos dados a partir da plataforma que alberga os dados (utilizando para isso as credências necessárias), de onde são extraídos ficheiros em formato .csv, que serão organizados em pastas de acordo com a estação e a Unidade de Recolha de Dados a que pertencem. De forma a normalizar os *data sets*, será atribuído um código de referência composto pela abreviatura do nome do projeto “USense”, seguido da estação a que os dados daquele ficheiro correspondem e da unidade e por fim a data de início de recolha dos dados, resultando por exemplo em “USense_24_de_Agosto_CO_25_06_2015”. Numa fase seguinte será realizada uma inventariação dos dados, para depois serem incorporados numa plataforma de descrição de dados - Dendro, onde serão descritos e por fim serão publicados num repositório de dados de investigação de acesso aberto.

DOCUMENTAÇÃO E METADADOS

Que documentação e metadados acompanhará os dados?

Numa primeira fase, durante a recolha dos dados estes não serão acompanhados de quaisquer anotações, tratando-se apenas de dados brutos. Depois de todos os dados serem recolhidos, serão utilizados alguns modelos de metadados mais gerais como é o caso do DC (Dublin Core) e da FoaF (Friend of a Friend) e outros de domínios mais específicos, como é o caso da SSN (Semantic Sensor Network), M3-Lite e DDI (Data Documentation Initiative) para descrever os dados recolhidos no projeto UrbanSense.
Os descritores selecionados pretendem descrever o projeto a um nível mais geral, como é o caso dos descritores "Creator"; "Description"; "Language"; "License"; "Publisher"; entre outros. Pretende-se ainda incluir uma descrição mais técnica, ao nível dos instrumentos utilizados na recolha dos dados, para isso serão utilizados descritores como, "Sensor"; "Measurement Range"; "Sensing Method Used". Por fim ao nível semântico, da análise do conteúdo do projeto e do período de recolha dos dados serão utilizados descritores como por exemplo: "Start Time"; "End Time"; "Start Date"; "End Date"; "Frequency"; "Methodology" e "Type".

ÉTICA E CONFORMIDADE LEGAL

De que forma se administra as questões éticas?

Não há problemas éticos relacionados com a recolha dos dados no caso do projeto UrbanSense.

Como irá gerir os direitos de autor e os direitos de propriedade intelectual?

Este projeto está a ser realizado em colaboração com vários parceiros. A equipa de investigação e desenvolvimento é composta por membros de várias universidades (Porto, Aveiro e Minho), numerosos institutos e grupos de I & D (Citta, IEETA, INESC TEC, TI, LabRP, Lepabe, LIACC e LSTS) e de diferentes domínios científicos (telecomunicações, eletrônica e engenharia da computação, psicologia e saúde, arquitetura e design, ciência da computação e engenharia química e ambiental.
O projeto conta ainda com a estreita cooperação industrial de várias empresas (pequenas, médias e grandes), bem como municípios da região do Porto, agências governamentais e outras instituições públicas.
Qualquer publicação relacionada com o projeto UrbanSense está dependente de autorização do responsável pelo projeto.

ARMAZENAMENTO E BACKUP

De que forma os dados serão armazenados durante a investigação?

Os dados recolhidos pelo UrbanSense são transmitidos pelas unidades estáticas de recolha de dados (DCU) para o Backoffice para processamento e armazenamento, através de WiFi APs ligados ao painel de fibra metropolitana ou por rede veicular Delay-Tolerant Networking (DTN). O servidor faz um backup regular dos dados.

Como serão geridos o acesso e a segurança?

Após a análise, organização e descrição dos dados, estes serão publicados num repositório de dados de investigação de acesso aberto e todos os acessos aos dados serão geridos a partir do servidor do repositório.

SELEÇÃO E PRESERVAÇÃO

Quais os dados que devem ser mantidos, partilhados e/ou preservados?

O conjunto de dados recolhidos no projeto, representam dados observacionais, recolhidos num determinado momento e local e por essa razão não se conseguem reproduzir em circunstâncias diferentes. Assim, todos os dados recolhidos são considerados de valor a longo prazo e serão mantidos e preservados, estando acessíveis através de um repositório de dados de investigação.

Qual o plano de preservação do *data set* a longo prazo?

Os dados depois de passar pelo processo de análise, organização e aplicação de metadados serão armazenados no repositório do Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência - INESC TEC, que permitirá aos utilizadores navegar e explorar os dados online e descarregar os *data sets* selecionados.
Os dados serão depositados no repositório de acordo com as políticas da instituição (INESC TEC), onde será preenchido um formulário de descrição acerca do projeto, dos descritores utilizados e dos ficheiros que integram o projeto.

PARTILHA DE DADOS

Como irá partilhar os dados?

Após a recolha, análise e organização dos dados serão selecionados e aplicados metadados aos dados para estes serem publicados.

Para além dos dados serem publicados num repositório de dados de investigação, também serão desenvolvidas ferramentas de visualização de dados, que permitirão que tanto investigadores como o público em geral tenham acesso aos dados.
Os metadados dos *data sets* mantidos no repositório institucional estarão disponíveis para pesquisa pelo público em geral.
A publicação dos dados permitirá a sua reutilização, uma vez que outras pessoas que demonstrem interesse na reutilização dos dados consigam agilizar processos noutras investigações, evitando a duplicação de trabalho.

Existe alguma restrição à partilha de dados?

Não haverá restrições na partilha dos dados. Após a conclusão do projeto, os dados serão disponibilizados em acesso aberto.

RESPONSABILIDADES E RECURSOS

Quem será responsável pela gestão dos dados?

O responsável pelo projeto será o responsável pela gestão dos dados. O Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência - INESC TEC será o responsável pela organização e armazenamento dos dados e pela gestão do repositório.

Que recursos serão necessários para o desenvolvimento do plano?

Os investigadores já possuem as infraestruturas necessárias para a implementação do Plano de Gestão de Dados.
A plataforma utilizada para aplicar o modelo de metadados para a descrição dos dados (Dendro) e o repositório de dados de investigação para a publicação dos *data set* (repositório de dados do INESC TEC), não terão qualquer custo para o projeto.

Anexo II. Inventário dos dados UrbanSense

Inventário dados UrbanSense							
ID	Nome do *dataset*	Estação			Tipo de dados	Tamanhodoficheiro	Nº registos
1	USense_24_de_Agosto_CO_2015_06_25	24 de Agosto	25-06-2015	04-02-2017	Dados observacionais (amostra)	2,78 MB	83491
2	USense_24_de_Agosto_Humidity_2015_06_25	24 de Agosto	25-06-2015	26-08-2017	Dados observacionais (amostra)	7,61 MB	252766
3	USense_24_de_Agosto_Luminosity_2015_06_25	24 de Agosto	25-06-2015	03-02-2017	Dados observacionais (amostra)	2,70 MB	81169
4	USense_24_de_Agosto_NO2_2015_06_25	24 de Agosto	25-06-2015	05-01-2017	Dados observacionais (amostra)	2,33 MB	70155
5	USense_24_de_Agosto_Noise_2015_07_29	24 de Agosto	29-07-2015	26-08-2017	Dados observacionais (amostra)	31 MB	1028687
6	USense_24_de_Agosto_O3_2015_06_25	24 de Agosto	25-06-2015	04-01-2017	Dados observacionais (amostra)	2,33 MB	70025
7	USense_24_de_Agosto_Particles1_2015_06_25	24 de Agosto	25-06-2015	26-08-2017	Dados observacionais (amostra)	8,34 MB	275641
8	USense_24_de_Agosto_Particles2_2015_06_25	24 de Agosto	25-06-2015	26-08-2017	Dados observacionais (amostra)	8,33 MB	275641
9	USense_24_de_Agosto_Precipitation_2015_06_25	24 de Agosto	25-06-2015	24-03-2017	Dados observacionais (amostra)	4,43 MB	149782
10	USense_24_de_Agosto_Solar radiation_2015_06_25	24 de Agosto	25-06-2015	04-02-2017	Dados observacionais (amostra)	4,01 MB	117179
11	USense_24_de_Agosto_Temperature_2015_06_25	24 de Agosto	25-06-2015	26-08-2017	Dados observacionais (amostra)	8,38 MB	275318
12	USense_24_de_Agosto_Wind direction_2015_06_25	24 de Agosto	25-06-2017	24-03-2017	Dados observacionais (amostra)	3,29 MB	112238
13	USense_24_de_Agosto_Wind speed_2015_06_25	24 de Agosto	25-06-2017	24-03-2017	Dados observacionais (amostra)	4,69 MB	158765
14	USense_Anemona_CO_2015_05_24	Anemona	24-05-2015	02-08-2015	Dados observacionais (amostra)	793 KB	23278
15	USense_Anemona_Humidity_2015_05_24	Anemona	24-05-2015	16-08-2015	Dados observacionais (amostra)	3,42 MB	112389
16	USense_Anemona_Luminosity_2015_05_24	Anemona	24-05-2015	29-07-2015	Dados observacionais (amostra)	2,84 MB	84159
17	USense_Anemona_NO2_2015_05_24	Anemona	24-05-2015	01-08-2015	Dados observacionais (amostra)	2,86 MB	85409
18	USense_Anemona_Noise_2015_05_24	Anemona	24-05-2015	30-07-2015	Dados observacionais (amostra)	110 KB	3481
19	USense_Anemona_O3_2015_05_24	Anemona	24-05-2015	01-08-2015	Dados observacionais (amostra)	2,86 MB	85717
20	USense_Anemona_Particles1_2015_05_24	Anemona	24-05-2015	16-08-2015	Dados observacionais (amostra)	3,48 MB	113669
21	USense_Anemona_Particles2_2015_05_24	Anemona	24-05-2015	16-08-2015	Dados observacionais (amostra)	3,58MB	113669
22	USense_Anemona_Precipitation_2015_05_24	Anemona	24-05-2015	26-05-2015	Dados observacionais (amostra)	80,9 KB	2670
23	Usense_Anemona_Solar radiation	Anemona			Dados observacionais (amostra)	80 bytes	0
24	USense_Anemona_Temperature_2015_05_24	Anemona	24-05-2015	16-08-2015	Dados observacionais (amostra)	3,46 MB	113666
25	USense_Anemona_Wind direction_2015_05_24	Anemona	24-05-2015	26-05-2015	Dados observacionais (amostra)	78,9 KB	2605
26	USense_Anemona_Wind speed_2015_05_24	Anemona	24-05-2015	26-05-2015	Dados observacionais (amostra)	84,8 KB	2801
27	USense_Av.Franca_CO_2015_06_18	Avenida de França	18-06-2015	18-11-2015	Dados observacionais (amostra)	3,43 MB	103122
28	USense_Av.Franca_Humidity_2015_06_18	Avenida de França	18-06-2015	03-04-2016	Dados observacionais (amostra)	2,90 MB	95348
29	USense_Av.Franca_Luminosity_2015_06_18	Avenida de França	18-06-2015	18-11-2015	Dados observacionais (amostra)	3,35 MB	100649
30	USense_Av.Franca_NO2_2015_06_18	Avenida de França	18-06-2015	18-11-2015	Dados observacionais (amostra)	3,42 MB	102900
31	USense_Av.Franca_Noise_2015_06_18	Avenida de França	18-06-2015	30-10-2015	Dados observacionais (amostra)	17,8 MB	586778
32	USense_Av.Franca_Noise_2015_10_31	Avenida de França	31-10-2015	09-11-2015	Dados observacionais (amostra)	16,8 MB	553135

33	USense_Av.Franca_Noise_2015_11_10	Avenida de França	10-11-2015	18-11-2015	Dados observacionais (amostra)	17,0 MB	559752
34	USense_Av.Franca_Noise_2016_03_02	Avenida de França	02-03-2016	22-03-2016	Dados observacionais (amostra)	17,9 MB	589483
35	USense_Av.Franca_O3_2015_06_18	Avenida de França	18-06-2015	18-11-2015	Dados observacionais (amostra)	3,37 MB	101269
36	USense_Av.Franca_Particles1_2015_06_18	Avenida de França	18-06-2015	03-04-2016	Dados observacionais (amostra)	3,14 MB	104514
37	USense_Av.Franca_Particles2_2015_06_18	Avenida de França	18-06-2015	03-04-2016	Dados observacionais (amostra)	3,26 MB	104514
38	USense_Av.Franca_Precipitation_2015_06_18	Avenida de França	18-06-2015	03-04-2016	Dados observacionais (amostra)	1,62 MB	55125
39	USense_Av.Franca_Solar radiation_2015_06_18	Avenida de França	18-06-2015	22-10-2015	Dados observacionais (amostra)	2,24 MB	65615
40	USense_Av.Franca_Temperature_2015_06_18	Avenida de França	18-06-2015	03-04-2016	Dados observacionais (amostra)	3,13 MB	102907
41	USense_Av.Franca_Wind direction_2015_06_18	Avenida de França	18-06-2015	22-10-2015	Dados observacionais (amostra)	1,34 MB	46298
42	USense_Av.Franca_Wind speed_2015_06_18	Avenida de França	18-06-2015	03-04-2016	Dados observacionais (amostra)	1,89 MB	64069
43	USense_Bolhao_pole1_CO_2015_06_26	Bolhão	26-06-2015	14-02-2017	Dados observacionais (amostra)	5,33 MB	160479
44	USense_Bolhao_pole1_Humidity_2015_06_26	Bolhão	26-06-2015	24-02-2017	Dados observacionais (amostra)	6,51 MB	214927
45	USense_Bolhao_pole1_Luminosity_2015_06_26	Bolhão	26-06-2015	24-02-2017	Dados observacionais (amostra)	7,32 MB	219388
46	USense_Bolhao_pole1_NO2_2015_06_26	Bolhão	26-06-2015	06-06-2016	Dados observacionais (amostra)	5,34 MB	160439
47	USense_Bolhao_pole1_Noise_2015_07_29	Bolhão	29-07-2015	07-10-2015	Dados observacionais (amostra)	14,4 MB	471880
48	USense_Bolhao_pole1_Noise_2016_03_01	Bolhão	01-03-2016	14-03-2016	Dados observacionais (amostra)	21,0 MB	688698
49	USense_Bolhao_pole1_Noise_2016_03_16	Bolhão	16-03-2016	04-04-2016	Dados observacionais (amostra)	30,2 MB	990580
50	USense_Bolhao_pole1_Noise_2016_05_16	Bolhão	16-05-2016	22-07-2016	Dados observacionais (amostra)	28,0 MB	918333
51	USense_Bolhao_pole1_Noise_2016_08_09	Bolhão	09-08-2016	29-09-2016	Dados observacionais (amostra)	17,6 MB	578254
52	USense_Bolhao_pole1_Noise_2016_09_30	Bolhão	30-09-2016	30-12-2016	Dados observacionais (amostra)	22,7 MB	744102
53	USense_Bolhao_pole1_Noise_2016_12_31	Bolhão	31-12-2016	24-02-2017	Dados observacionais (amostra)	17,2 MB	564040
54	USense_Bolhao_pole1_O3_2015_06_26	Bolhão	26-06-2015	06-06-2016	Dados observacionais (amostra)	5,33 MB	160083
55	USense_Bolhao_pole1_Particles1_2015_06_26	Bolhão	26-06-2015	24-02-2017	Dados observacionais (amostra)	6,92 MB	224764
56	USense_Bolhao_pole1_Particles2_2015_06_26	Bolhão	26-06-2015	24-02-2017	Dados observacionais (amostra)	7,02 MB	224764
57	USense_Bolhao_pole1_Temperature_2015_06_26	Bolhão	26-06-2015	24-02-2017	Dados observacionais (amostra)	6,80 MB	223260
58	USense_Campo_Alegre_CO_2015_07_17	Campo Alegre	17-07-2015	31-05-2016	Dados observacionais (amostra)	3,39 MB	101885
59	USense_Campo_Alegre_Humidity_2015_07_17	Campo Alegre	17-07-2015	16-09-2015	Dados observacionais (amostra)	2,61 MB	85812
60	USense_Campo_Alegre_Luminosity_2015_07_17	Campo Alegre	17-07-2015	01-06-2016	Dados observacionais (amostra)	3,36 MB	101014
61	USense_Campo_Alegre_NO2_2015_07_17	Campo Alegre	17-07-2015	30-05-2016	Dados observacionais (amostra)	3,30 MB	99199
62	USense_Campo_Alegre_Noise_2015_07_29	Campo Alegre	29-07-2015	06-03-2016	Dados observacionais (amostra)	7,49 MB	244334
63	USense_Campo_Alegre_Noise_2016_03_10	Campo Alegre	10-03-2016	30-03-2016	Dados observacionais (amostra)	27,3 MB	885610
64	USense_Campo_Alegre_Noise_2016_03_31	Campo Alegre	31-03-2016	30-05-2016	Dados observacionais (amostra)	26,3 MB	859372
65	USense_Campo_Alegre_Noise_2016_05_31	Campo Alegre	31-05-2016	28-11-2016	Dados observacionais (amostra)	2,97 MB	96341
66	USense_Campo_Alegre_O3_2015_07_17	Campo Alegre	17-07-2015	31-05-2016	Dados observacionais (amostra)	3,32 MB	100052
67	USense_Campo_Alegre_Particles1_2015_07_17	Campo Alegre	17-07-2015	29-11-2016	Dados observacionais (amostra)	3,09 MB	102580
68	USense_Campo_Alegre_Particles2_2015_07_17	Campo Alegre	17-07-2015	29-11-2016	Dados observacionais (amostra)	3,21 MB	102580
69	USense_Campo_Alegre_Temperature_2015_07_17	Campo Alegre	17-07-2015	16-09-2015	Dados observacionais (amostra)	2,62 MB	85874
70	USense_Candido_dos_Reis_CO_2015_06_19	Cândido dos Reis	19-06-2015	02-06-2016	Dados observacionais (amostra)	3,88 MB	116646
71	USense_Candido_dos_Reis_Humidity_2015_06_19	Cândido dos Reis	19-06-2015	02-06-2016	Dados observacionais (amostra)	3,54 MB	116053

72	USense_Candido_dos_Reis_Luminosity_2015_06_19	Cândido dos Reis	19-06-2015	02-06-2016	Dados observacionais (amostra)	3,77 MB	113489
73	USense_Candido_dos_Reis_NO2_2015_06_19	Cândido dos Reis	19-06-2015	02-06-2016	Dados observacionais (amostra)	3,73 MB	112303
74	USense_Candido_dos_Reis_Noise_2015_07_29	Cândido dos Reis	29-07-2015	29-07-2015	Dados observacionais (amostra)	269 bytes	6
75	USense_Candido_dos_Reis_O3_2015_06_19	Cândido dos Reis	19-06-2015	02-06-2016	Dados observacionais (amostra)	3,73 MB	112264
76	USense_Candido_dos_Reis_Particles1_2015_06_19	Cândido dos Reis	19-06-2015	02-06-2016	Dados observacionais (amostra)	3,39 MB	112330
77	USense_Candido_dos_Reis_Particles2_2015_06_19	Cândido dos Reis	19-06-2015	02-06-2016	Dados observacionais (amostra)	3,53 MB	112330
78	USense_Candido_dos_Reis_Precipitation_2015_06_19	Cândido dos Reis	19-06-2015	02-06-2016	Dados observacionais (amostra)	1,17 MB	39761
79	USense_Candido_dos_Reis_Solar radiation_2015_06_19	Cândido dos Reis	19-06-2015	02-06-2016	Dados observacionais (amostra)	1,81 MB	52987
80	USense_Candido_dos_Reis_Temperature_2015_06_19	Cândido dos Reis	19-06-2015	02-06-2016	Dados observacionais (amostra)	3,55 MB	116448
81	USense_Candido_dos_Reis_Wind direction_2015_06_19	Cândido dos Reis	19-06-2015	20-08-2015	Dados observacionais (amostra)	987 KB	32975
82	USense_Candido_dos_Reis_Wind speed_2015_06_19	Cândido dos Reis	19-06-2015	02-06-2016	Dados observacionais (amostra)	1,53 MB	51946
83	USense_Casa_da_Musica_CO_2015_07_14	Casa da Música	14-07-2015	12-05-2017	Dados observacionais (amostra)	9,56 MB	287432
84	USense_Casa_da_Musica_Humidity_2015_07_14	Casa da Música	14-07-2015	26-08-2017	Dados observacionais (amostra)	9,01 MB	297125
85	USense_Casa_da_Musica_Luminosity_2015_07_14	Casa da Música	14-07-2015	29-10-2015	Dados observacionais (amostra)	4,36 MB	129690
86	USense_Casa_da_Musica_NO2_2015_07_14	Casa da Música	14-07-2015	23-05-2017	Dados observacionais (amostra)	7,71 MB	231174
87	USense_Casa_da_Musica_Noise_2015_07_29	Casa da Música	29-07-2015	09-10-2015	Dados observacionais (amostra)	21,5 MB	706175
88	USense_Casa_da_Musica_Noise_2016_01_28	Casa da Música	28-01-2016	19-06-2016	Dados observacionais (amostra)	22,2 MB	729031
89	USense_Casa_da_Musica_Noise_2016_07_17	Casa da Música	17-07-2016	30-10-2016	Dados observacionais (amostra)	30,1 MB	977197
90	USense_Casa_da_Musica_Noise_2016_10_31	Casa da Música	31-10-2016	23-12-2016	Dados observacionais (amostra)	29,3 MB	963000
91	USense_Casa_da_Musica_Noise_2016_12_24	Casa da Música	24-12-2016	27-02-2017	Dados observacionais (amostra)	29,2 MB	957065
92	USense_Casa_da_Musica_Noise_2017_02_28	Casa da Música	28-02-2017	09-06-2017	Dados observacionais (amostra)	31,3 MB	1023434
93	USense_Casa_da_Musica_Noise_2017_06_10	Casa da Música	10-06-2017	29-06-2017	Dados observacionais (amostra)	31,3 MB	1013999
94	USense_Casa_da_Musica_Noise_2017_06_30	Casa da Música	30-06-2017	16-07-2017	Dados observacionais (amostra)	28,7 MB	941785
95	USense_Casa_da_Musica_Noise_2017_07_17	Casa da Música	17-07-2017	29-07-2017	Dados observacionais (amostra)	30,3 MB	994849
96	USense_Casa_da_Musica_Noise_2017_07_30	Casa da Música	30-07-2017	12-08-2017	Dados observacionais (amostra)	31,2 MB	1020651
97	USense_Casa_da_Musica_Noise_2017_08_13	Casa da Música	13-08-2017	19-08-2017	Dados observacionais (amostra)	18,3 MB	600748
98	USense_Casa_da_Musica_Noise_2017_08_20	Casa da Música	20-08-2017	26-08-2017	Dados observacionais (amostra)	18,0 MB	591608
99	USense_Casa_da_Musica_O3_2015_07_14	Casa da Música	14-07-2015	11-05-2017	Dados observacionais (amostra)	7,67 MB	229701
100	USense_Casa_da_Musica_Particles1_2015_07_14	Casa da Música	14-07-2015	26-08-2017	Dados observacionais (amostra)	9,38 MB	306764
101	USense_Casa_da_Musica_Particles2_2015_07_14	Casa da Música	14-07-2015	26-08-2017	Dados observacionais (amostra)	9,52 MB	306764
102	USense_Casa_da_Musica_Temperature_2015_07_14	Casa da Música	14-07-2015	26-08-2017	Dados observacionais (amostra)	9,26 MB	304265
103	USense_Castelo_do_Queijo_CO_2015_06_05	Castelo do Queijo	05-06-2015	21-08-2015	Dados observacionais (amostra)	2,70 MB	81180
104	USense_Castelo_do_Queijo_Humidity_2015_06_05	Castelo do Queijo	05-06-2015	21-08-2015	Dados observacionais (amostra)	3,01 MB	99976
105	USense_Castelo_do_Queijo_Luminosity_2015_06_05	Castelo do Queijo	05-06-2015	21-08-2015	Dados observacionais (amostra)	2,73 MB	81289
106	USense_Castelo_do_Queijo_NO2_2015_06_05	Castelo do Queijo	05-06-2015	21-08-2015	Dados observacionais (amostra)	3,41 MB	102419
107	USense_Castelo_do_Queijo_Noise_2015_07_29	Castelo do Queijo	29-07-2015	21-08-2015	Dados observacionais (amostra)	877 KB	28082
108	USense_Castelo_do_Queijo_O3_2015_06_09	Castelo do Queijo	09-06-2015	21-08-2015	Dados observacionais (amostra)	3,40 MB	102385
109	USense_Castelo_do_Queijo_Particles1_2015_06_09	Castelo do Queijo	09-06-2015	21-08-2015	Dados observacionais (amostra)	3,12 MB	102399
110	USense_Castelo_do_Queijo_Particles2_2015_06_09	Castelo do Queijo	09-06-2015	21-08-2015	Dados observacionais (amostra)	3,25 MB	102399

111	USense_Castelo_do_Queijo_Precipitation_2015_06_05	Castelo do Queijo	05-06-2015	21-08-2015	Dados observacionais (amostra)	2,42 MB	82067
112	USense_Castelo_do_Queijo_Sola radiation_2015_06_12	Castelo do Queijo	12-06-2015	21-08-2015	Dados observacionais (amostra)	3,17 MB	93533
113	USense_Castelo_do_Queijo_Temperature_2015_06_05	Castelo do Queijo	05-06-2015	21-08-2015	Dados observacionais (amostra)	3,12 MB	102308
114	USense_Castelo_do_Queijo_Wind direction_2015_06_05	Castelo do Queijo	05-06-2015	21-08-2015	Dados observacionais (amostra)	2,19 MB	74969
115	USense_Castelo_do_Queijo_Wind speed_2015_06_05	Castelo do Queijo	05-06-2015	21-08-2015	Dados observacionais (amostra)	2,50 MB	84198
116	USense_Combatentes_CO_2015_05_26	Combatentes	26-05-2015	17-06-2017	Dados observacionais (amostra)	7,06 MB	212256
117	USense_Combatentes_Humidity_2015_05_26	Combatentes	26-05-2015	26-11-2016	Dados observacionais (amostra)	6,61 MB	216775
118	USense_Combatentes_Luminosity_2015_05_26	Combatentes	26-05-2015	19-06-2017	Dados observacionais (amostra)	7,08 MB	213439
119	USense_Combatentes_NO2_2015_05_26	Combatentes	26-05-2015	21-06-2017	Dados observacionais (amostra)	7,05 MB	211871
120	USense_Combatentes_Noise_2015_07_29	Combatentes	29-07-2015	07-10-2015	Dados observacionais (amostra)	20,8 MB	683858
121	USense_Combatentes_Noise_2016_01_29	Combatentes	29-01-2016	06-03-2016	Dados observacionais (amostra)	5,66 MB	185742
122	USense_Combatentes_Noise_2016_03_07	Combatentes	07-03-2016	24-03-2016	Dados observacionais (amostra)	26,9 MB	882091
123	USense_Combatentes_Noise_2016_03_25	Combatentes	25-03-2016	29-09-2016	Dados observacionais (amostra)	12,9 MB	425159
124	USense_Combatentes_Noise_2016_09_30	Combatentes	30-09-2016	26-11-2016	Dados observacionais (amostra)	17,8 MB	585201
125	USense_Combatentes_Noise_2017_04_13	Combatentes	13-04-2017	08-06-2017	Dados observacionais (amostra)	17,8 MB	583667
126	USense_Combatentes_Noise_2017_06_09	Combatentes	09-06-2017	22-06-2017	Dados observacionais (amostra)	26,3 MB	862136
127	USense_Combatentes_O3_2015_05_26	Combatentes	26-05-2015	17-06-2017	Dados observacionais (amostra)	7,06 MB	212256
128	USense_Combatentes_Particles1_2015_05_26	Combatentes	26-05-2015	22-06-2017	Dados observacionais (amostra)	6,90 MB	229247
129	USense_Combatentes_Particles2_2015_05_26	Combatentes	26-05-2015	22-06-2017	Dados observacionais (amostra)	7,16 MB	229247
130	USense_Combatentes_Temperature_2015_05_26	Combatentes	26-05-2015	26-11-2016	Dados observacionais (amostra)	6,65 MB	218345
131	USense_Congregados_CO_2015_10_01	Congregados	01-10-2015	22-10-2015	Dados observacionais (amostra)	999 KB	29044
132	USense_Congregados_Humidity_2015_10_01	Congregados	01-10-2015	16-10-2015	Dados observacionais (amostra)	551 KB	17648
133	USense_Congregados_Luminosity_2015_10_01	Congregados	01-10-2015	16-10-2015	Dados observacionais (amostra)	729 KB	20757
134	USense_Congregados_NO2_2015_10_01	Congregados	01-10-2015	22-10-2015	Dados observacionais (amostra)	975 KB	28639
135	USense_Congregados_Noise	Congregados			Dados observacionais (amostra)	72 bytes	0
136	USense_Congregados_O3_2015_10_01	Congregados	01-10-2015	22-10-2015	Dados observacionais (amostra)	989 KB	29027
137	USense_Congregados_Particles1_2015_10_01	Congregados	01-10-2015	22-10-2015	Dados observacionais (amostra)	889 KB	29043
138	USense_Congregados_Particles2_2015_10_01	Congregados	01-10-2015	22-10-2015	Dados observacionais (amostra)	927 KB	29043
139	USense_Congregados_Precipitation_2015_10_01	Congregados	01-10-2015	22-10-2015	Dados observacionais (amostra)	779 KB	25653
140	USense_Congregados_Solar radiation_2015_10_01	Congregados	01-10-2015	22-10-2015	Dados observacionais (amostra)	449 KB	13778
141	USense_Congregados_Temperature_2015_10_01	Congregados	01-10-2015	16-10-2015	Dados observacionais (amostra)	621 KB	19891
142	USense_Congregados_Wind direction_2015_10_01	Congregados	01-10-2015	22-10-2015	Dados observacionais (amostra)	314 KB	10502
143	USense_Congregados_Wind speed_2015_10_01	Congregados	01-10-2015	22-10-2015	Dados observacionais (amostra)	882 KB	29045
144	USense_D.Manuel II_CO_2015_06_18	D. Manuel II	18-06-2015	25-03-2017	Dados observacionais (amostra)	7,71 MB	227732
145	USense_D.Manuel II_Humidity_2015_06_18	D. Manuel II	18-06-2015	25-03-2017	Dados observacionais (amostra)	6,44 MB	213262
146	USense_D.Manuel II_Luminosity_2015_06_18	D. Manuel II	18-06-2015	10-08-2016	Dados observacionais (amostra)	4,94 MB	148597
147	USense_D.Manuel II_NO2_2015_06_18	D. Manuel II	18-06-2015	25-03-2017	Dados observacionais (amostra)	7,74 MB	228403
148	USense_D.Manuel II_Noise_2015_06_18	D. Manuel II	18-06-2015	04-11-2015	Dados observacionais (amostra)	27,9 MB	915267
149	USense_D.Manuel II_Noise_2015_11_05	D. Manuel II	05-11-2015	18-03-2016	Dados observacionais (amostra)	28,2 MB	919867

150	USense_D.Manuel II_Noise_2016_03_19	D. Manuel II	19-03-2016	25-10-2016	Dados observacionais (amostra)	25,8 MB	843447
151	USense_D.Manuel II_Noise_2016_10_25	D. Manuel II	25-10-2016	26-11-2016	Dados observacionais (amostra)	22,8 MB	748055
152	USense_D.Manuel II_Noise_2016_11_28	D. Manuel II	28-11-2016	06-01-2017	Dados observacionais (amostra)	28,3 MB	924524
153	USense_D.Manuel II_Noise_2017_01_06	D. Manuel II	06-01-2017	25-03-2017	Dados observacionais (amostra)	8,50 MB	278579
154	USense_D.Manuel II_O3_2015_06_18	D. Manuel II	18-06-2015	25-03-2017	Dados observacionais (amostra)	7,26 MB	214776
155	USense_D.Manuel II_Particles1_2015_06_18	D. Manuel II	18-06-2015	25-03-2017	Dados observacionais (amostra)	7,08 MB	238509
156	USense_D.Manuel II_Particles2_2015_06_18	D. Manuel II	18-06-2015	25-03-2017	Dados observacionais (amostra)	7,11 MB	238509
157	USense_D.Manuel II_Temperature_2015_06_18	D. Manuel II	18-06-2015	25-03-2017	Dados observacionais (amostra)	7,06 MB	231800
158	USense_Damiao_de_gois_CO_2015_06_23	Damião de Góis	23-06-2015	21-10-2015	Dados observacionais (amostra)	4,45 MB	133853
159	USense_Damiao_de_gois_Humidity_2014_08_06	Damião de Góis	06-08-2014	21-10-2015	Dados observacionais (amostra)	8,70 MB	285982
160	USense_Damiao_de_gois_Luminosity_2015_06_23	Damião de Góis	23-06-2015	21-10-2015	Dados observacionais (amostra)	5,37 MB	160909
161	USense_Damiao_de_gois_NO2_2015_06_23	Damião de Góis	23-06-2015	21-10-2015	Dados observacionais (amostra)	5,36 MB	161261
162	Usense_Damiao_de_gois_Noise_2014_08_06	Damião de Góis	06-08-2014	09-10-2015	Dados observacionais (amostra)	25,1 MB	824219
163	Usense_Damiao_de_gois_Noise_2015_10_10	Damião de Góis	10-10-2015	21-10-2015	Dados observacionais (amostra)	23,8 MB	781942
164	USense_Damiao_de_gois_O3_2015_06_23	Damião de Góis	23-06-2015	21-10-2015	Dados observacionais (amostra)	5,36 MB	161289
165	USense_Damiao_de_gois_Particles1_2015_06_23	Damião de Góis	23-06-2015	21-10-2015	Dados observacionais (amostra)	4,86 MB	161293
166	USense_Damiao_de_gois_Particles2_2015_06_23	Damião de Góis	23-06-2015	21-10-2015	Dados observacionais (amostra)	5,06 MB	161293
167	USense_Damiao_de_gois_Precipitation_2014_08_06	Damião de Góis	06-08-2014	21-10-2015	Dados observacionais (amostra)	9,31 MB	314794
168	USense_Damiao_de_gois_Solar radiation_2014_08_06	Damião de Góis	06-08-2014	21-10-2015	Dados observacionais (amostra)	10,5 MB	308192
169	USense_Damiao_de_gois_Temperature_2014_08_06	Damião de Góis	06-08-2014	21-10-2015	Dados observacionais (amostra)	8,76 MB	287246
170	USense_Damiao_de_gois_Wind direction_2014_08_06	Damião de Góis	06-08-2014	21-10-2015	Dados observacionais (amostra)	8,10 MB	277289
171	USense_Damiao_de_gois_Wind speed_2014_08_06	Damião de Góis	06-08-2014	21-10-2015	Dados observacionais (amostra)	9,65 MB	326457
172	USense_Estadio_do_Dragao_CO_2015_07_10	Estádio do Dragão	10-07-2015	10-12-2015	Dados observacionais (amostra)	6,17 MB	185485
173	USense_Estadio_do_Dragao_Humidity_2015_07_10	Estádio do Dragão	10-07-2015	10-12-2015	Dados observacionais (amostra)	5,56 MB	183262
174	USense_Estadio_do_Dragao_Luminosity_2015_07_10	Estádio do Dragão	10-07-2015	10-12-2015	Dados observacionais (amostra)	6,14 MB	183804
175	USense_Estadio_do_Dragao_NO2_2015_07_10	Estádio do Dragão	10-07-2015	10-12-2015	Dados observacionais (amostra)	6,17 MB	185495
176	USense_Estadio_do_Dragao_Noise_2015_08_03	Estádio do Dragão	03-08-2015	09-10-2015	Dados observacionais (amostra)	24,1 MB	790375
177	USense_Estadio_do_Dragao_Noise_2015_07_10	Estádio do Dragão	10-10-2015	17-10-2015	Dados observacionais (amostra)	15,4 MB	505731
178	USense_Estadio_do_Dragao_O3_2015_07_10	Estádio do Dragão	10-07-2015	10-12-2015	Dados observacionais (amostra)	6,16 MB	185049
179	USense_Estadio_do_Dragao_Particles1_2015_07_10	Estádio do Dragão	10-07-2015	10-12-2015	Dados observacionais (amostra)	5,58 MB	185496
180	USense_Estadio_do_Dragao_Particles2_2015_07_10	Estádio do Dragão	10-07-2015	10-12-2015	Dados observacionais (amostra)	5,85 MB	185496
181	USense_Estadio_do_Dragao_Temperature_2015_07_10	Estádio do Dragão	10-07-2015	10-12-2015	Dados observacionais (amostra)	5,64 MB	185237
182	USense_FEUP_P1_CO_2015_05_18	FEUP	18-05-2015	29-12-2015	Dados observacionais (amostra)	4,03 MB	121173
183	USense_FEUP_P1_Humidity_2015_05_18	FEUP	18-05-2015	13-01-2017	Dados observacionais (amostra)	3,80 MB	125290
184	USense_FEUP_P1_Luminosity_2015_05_18	FEUP	18-05-2015	29-12-2015	Dados observacionais (amostra)	4,05 MB	120300
185	USense_FEUP_P1_NO2_2015_05_18	FEUP	18-05-2015	29-12-2015	Dados observacionais (amostra)	2,46 MB	74122
186	USense_FEUP_P1_Noise_2016_06_16	FEUP	16-06-2016	16-12-2016	Dados observacionais (amostra)	5,44 MB	184141
187	USense_FEUP_P1_O3_2015_05_18	FEUP	18-05-2015	29-12-2015	Dados observacionais (amostra)	3,99 MB	119539
188	USense_FEUP_P1_Particles1_2015_05_18	FEUP	18-05-2015	16-12-2016	Dados observacionais (amostra)	3,88 MB	126801

189	USense_FEUP_P1_Particles2_2015_05_18	FEUP	18-05-2015	16-12-2016	Dados observacionais (amostra)	4,05 MB	126801
190	USense_FEUP_P1_Temperature_2015_05_18	FEUP	18-05-2015	16-12-2016	Dados observacionais (amostra)	3,86 MB	126649
191	USense_Fundacao_de_Serralves_CO_2015_07_28	Fundação de Serralves	28-07-2015	30-09-2016	Dados observacionais (amostra)	8,45 MB	253351
192	USense_Fundacao_de_Serralves_Humidity_2015_07_28	Fundação de Serralves	28-07-2015	23-10-2016	Dados observacionais (amostra)	5,51 MB	184840
193	USense_Fundacao_de_Serralves_Luminosity_2015_07_28	Fundação de Serralves	28-07-2015	23-10-2016	Dados observacionais (amostra)	2,68 MB	78926
194	USense_Fundacao_de_Serralves_NO2_2015_07_28	Fundação de Serralves	28-07-2015	30-09-2016	Dados observacionais (amostra)	8,50 MB	255343
195	USense_Fundacao_de_Serralves_Noise_2015_09_29	Fundação de Serralves	29-09-2015	30-10-2015	Dados observacionais (amostra)	21,6 MB	710460
196	USense_Fundacao_de_Serralves_Noise_2015_10_31	Fundação de Serralves	31-10-2015	09-11-2015	Dados observacionais (amostra)	26,3 MB	863628
197	USense_Fundacao_de_Serralves_Noise_2015_11_10	Fundação de Serralves	10-11-2015	19-11-2015	Dados observacionais (amostra)	26,3 MB	863581
198	USense_Fundacao_de_Serralves_Noise_2015_11_20	Fundação de Serralves	20-11-2015	29-11-2015	Dados observacionais (amostra)	26,3 MB	863627
199	USense_Fundacao_de_Serralves_Noise_2015_11_30	Fundação de Serralves	30-11-2015	09-12-2015	Dados observacionais (amostra)	26,3 MB	863591
200	USense_Fundacao_de_Serralves_Noise_2015_12_10	Fundação de Serralves	10-12-2015	19-12-2015	Dados observacionais (amostra)	26,2 MB	861790
201	USense_Fundacao_de_Serralves_Noise_2015_12_20	Fundação de Serralves	20-12-2015	30-12-2015	Dados observacionais (amostra)	28,9 MB	949848
202	USense_Fundacao_de_Serralves_Noise_2015_12_31	Fundação de Serralves	31-12-2015	08-01-2016	Dados observacionais (amostra)	23,7 MB	777105
203	USense_Fundacao_de_Serralves_Noise_2016_01_09	Fundação de Serralves	09-01-2016	19-01-2016	Dados observacionais (amostra)	28,9 MB	949701
204	USense_Fundacao_de_Serralves_Noise_2016_01_20	Fundação de Serralves	20-01-2016	02-02-2016	Dados observacionais (amostra)	24,9 MB	816327
205	USense_Fundacao_de_Serralves_Noise_2016_02_03	Fundação de Serralves	03-02-2016	28-02-2016	Dados observacionais (amostra)	22,0 MB	723005
206	USense_Fundacao_de_Serralves_Noise_2016_02_29	Fundação de Serralves	29-02-2016	14-03-2016	Dados observacionais (amostra)	19,7 MB	647057
207	USense_Fundacao_de_Serralves_Noise_2016_03_15	Fundação de Serralves	15-03-2016	09-04-2016	Dados observacionais (amostra)	25,4 MB	832520
208	USense_Fundacao_de_Serralves_Noise_2016_04_10	Fundação de Serralves	10-04-2016	29-04-2016	Dados observacionais (amostra)	26,1 MB	858495
209	USense_Fundacao_de_Serralves_Noise_2016_04_30	Fundação de Serralves	30-04-2016	19-05-2016	Dados observacionais (amostra)	26,0 MB	855089
210	USense_Fundacao_de_Serralves_Noise_2016_05_20	Fundação de Serralves	20-05-2016	12-06-2016	Dados observacionais (amostra)	31,6 MB	1036205
211	USense_Fundacao_de_Serralves_Noise_2016_06_13	Fundação de Serralves	13-06-2016	14-08-2016	Dados observacionais (amostra)	26,3 MB	861875
212	USense_Fundacao_de_Serralves_Noise_2016_08_15	Fundação de Serralves	15-08-2016	26-08-2016	Dados observacionais (amostra)	9,11 MB	298725
213	USense_Fundacao_de_Serralves_O3_2015_07_28	Fundação de Serralves	28-07-2015	30-09-2016	Dados observacionais (amostra)	7,94 MB	238691
214	USense_Fundacao_de_Serralves_Particles1_2015_07_28	Fundação de Serralves	28-07-2015	23-10-2016	Dados observacionais (amostra)	7,74 MB	257301
215	USense_Fundacao_de_Serralves_Particles2_2015_07_28	Fundação de Serralves	28-07-2015	23-10-2016	Dados observacionais (amostra)	8,04 MB	257301
216	USense_Fundacao_de_Serralves_Precipitation_2015_07_28	Fundação de Serralves	28-07-2015	23-10-2016	Dados observacionais (amostra)	4,04 MB	136719
217	USense_Fundacao_de_Serralves_Solar radiation_2015_07_28	Fundação de Serralves	28-07-2015	23-10-2016	Dados observacionais (amostra)	4,68 MB	138458
218	USense_Fundacao_de_Serralves_Temperature_2015_07_28	Fundação de Serralves	28-07-2015	23-10-2016	Dados observacionais (amostra)	7,42 MB	243626
219	USense_Fundacao_de_Serralves_Wind direction_2015_07_28	Fundação de Serralves	28-07-2015	30-09-2016	Dados observacionais (amostra)	3,49 MB	120599
220	USense_Fundacao_de_Serralves_Wind speed_2015_07_28	Fundação de Serralves	28-07-2015	23-10-2016	Dados observacionais (amostra)	3,90 MB	132119
221	USense_Hospital_de_S.Joao_CO_2015_07_20	Hospital de S.João	20-07-2015	04-04-2016	Dados observacionais (amostra)	1,18 MB	35640
222	USense_Hospital_de_S.Joao_Humidity_2015_07_20	Hospital de S.João	20-07-2015	18-10-2016	Dados observacionais (amostra)	6,18 MB	205634
223	USense_Hospital_de_S.Joao_Luminosity_2015_07_20	Hospital de S.João	20-07-2015	28-03-2016	Dados observacionais (amostra)	1,15 MB	34594
224	USense_Hospital_de_S.Joao_NO2_2015_07_20	Hospital de S.João	20-07-2015	01-04-2016	Dados observacionais (amostra)	1,18 MB	35656
225	USense_Hospital_de_S.Joao_Noise_2015_07_29	Hospital de S.João	29-07-2015	09-03-2016	Dados observacionais (amostra)	14,1 MB	462207
226	Usense_Hospital_de_S.Joao_Noise_2016_03_10	Hospital de S.João	10-03-2016	30-03-2016	Dados observacionais (amostra)	27,7 MB	904578
227	USense_Hospital_de_S.Joao_Noise_2016_03_31	Hospital de S.João	31-03-2016	18-10-2016	Dados observacionais (amostra)	26,4 MB	862215

228	USense_Hospital_de_S.Joao_O3_2015_07_20	Hospital de S.João	20-07-2015	01-04-2016	Dados observacionais (amostra)	1,18 MB	35484
229	USense_Hospital_de_S.Joao_Particles1_2015_07_20	Hospital de S.João	20-07-2015	18-10-2016	Dados observacionais (amostra)	7,07 MB	238432
230	USense_Hospital_de_S.Joao_Particles2_2015_07_20	Hospital de S.João	20-07-2015	18-10-2016	Dados observacionais (amostra)	7,11 MB	238432
231	USense_Hospital_de_S.Joao_Temperature_2015_07_20	Hospital de S.João	20-07-2015	18-10-2016	Dados observacionais (amostra)	7,23 MB	237042
232	USense_Pr_Liberdade_Cardosas_CO_2015_07_03	Praça da Liberdade- Cardosas	03-07-2015	16-06-2016	Dados observacionais (amostra)	6,99 MB	209928
233	USense_Pr_Liberdade_Cardosas_Humidity_2015_07_03	Praça da Liberdade- Cardosas	03-07-2015	21-11-2016	Dados observacionais (amostra)	6,42 MB	210720
234	USense_Pr_Liberdade_Cardosas_Luminosity_2015_07_03	Praça da Liberdade- Cardosas	03-07-2015	17-05-2016	Dados observacionais (amostra)	6,81 MB	204844
235	USense_Pr_Liberdade_Cardosas_NO2_2015_07_03	Praça da Liberdade- Cardosas	03-07-2015	16-06-2016	Dados observacionais (amostra)	7,00 MB	209808
236	USense_Pr_Liberdade_Cardosas_Noise_2015_07_29	Praça da Liberdade- Cardosas	29-07-2015	29-01-2016	Dados observacionais (amostra)	25,5 MB	836686
237	USense_Pr_Liberdade_Cardosas_Noise_2016_03_02	Praça da Liberdade- Cardosas	02-03-2016	14-03-2016	Dados observacionais (amostra)	15,7 MB	504381
238	USense_Pr_Liberdade_Cardosas_Noise_2016_03_15	Praça da Liberdade- Cardosas	15-03-2016	14-04-2017	Dados observacionais (amostra)	23,9 MB	772517
239	USense_Pr_Liberdade_Cardosas_O3_2015_07_03	Praça da Liberdade- Cardosas	03-07-2015	16-06-2016	Dados observacionais (amostra)	6,94 MB	208624
240	USense_Pr_Liberdade_Cardosas_Particles1_2015_07_03	Praça da Liberdade- Cardosas	03-07-2015	14-04-2017	Dados observacionais (amostra)	6,45 MB	212950
241	USense_Pr_Liberdade_Cardosas_Particles2_2015_07_03	Praça da Liberdade- Cardosas	03-07-2015	14-04-2017	Dados observacionais (amostra)	6,78 MB	212950
242	USense_Pr_Liberdade_Cardosas_Temperature_2015_07_03	Praça da Liberdade- Cardosas	03-07-2015	21-11-2016	Dados observacionais (amostra)	6,49 MB	212807
243	Usense_Praca_da_Galiza_CO_2015_06_18	Praça da Galiza	18-06-2015	18-11-2016	Dados observacionais (amostra)	8,29 MB	249299
244	Usense_Praca_da_Galiza_Humidity_2015_06_18	Praça da Galiza	18-06-2015	18-08-2017	Dados observacionais (amostra)	7,69 MB	252543
245	Usense_Praca_da_Galiza_Luminosity_2015_06_18	Praça da Galiza	18-06-2015	19-03-2017	Dados observacionais (amostra)	8,27 MB	248513
246	Usense_Praca_da_Galiza_NO2_2015_06_18	Praça da Galiza	18-06-2015	18-11-2016	Dados observacionais (amostra)	8,27 MB	248558
247	USense_Praca_da_Galiza_Noise_2015_06_18	Praça da Galiza	18-06-2015	27-09-2015	Dados observacionais (amostra)	10,1 MB	333347
248	USense_Praca_da_Galiza_Noise_2016_01_29	Praça da Galiza	29-01-2016	18-11-2016	Dados observacionais (amostra)	23,9 MB	784551
249	Usense_Praca_da_Galiza_O3_2015_06_18	Praça da Galiza	18-06-2015	18-11-2016	Dados observacionais (amostra)	8,24 MB	246959
250	Usense_Praca_da_Galiza_Particles1_2015_06_18	Praça da Galiza	18-06-2015	18-08-2017	Dados observacionais (amostra)	7,64 MB	253062
251	Usense_Praca_da_Galiza_Particles2_2015_06_18	Praça da Galiza	18-06-2015	18-08-2017	Dados observacionais (amostra)	7,97 MB	253062
252	Usense_Praca_da_Galiza_Temperature_2015_06_18	Praça da Galiza	18-06-2015	18-08-2017	Dados observacionais (amostra)	7,71 MB	252930
253	USense_Praca_do_Marques_CO_2015_05_24	Praça do Marquês- pole1	24-05-2015	02-06-2016	Dados observacionais (amostra)	10,9 MB	328917
254	USense_Praca_do_Marques_Humidity_2015_05_24	Praça do Marquês- pole1	24-05-2015	02-06-2016	Dados observacionais (amostra)	8,89 MB	291571
255	USense_Praca_do_Marques_Luminosity_2015_05_24	Praça do Marquês- pole1	24-05-2015	02-06-2016	Dados observacionais (amostra)	11,1 MB	328993
256	USense_Praca_do_Marques_NO2_2015_05_24	Praça do Marquês- pole1	24-05-2015	02-06-2016	Dados observacionais (amostra)	10,9 MB	329040
257	USense_Praca_do_Marques_Noise_2015_07_29	Praça do Marquês- pole1	29-07-2015	30-10-2015	Dados observacionais (amostra)	24,1 MB	790909
258	USense_Praca_do_Marques_Noise_2015_10_31	Praça do Marquês- pole1	31-10-2015	09-11-2015	Dados observacionais (amostra)	26,0 MB	853748
259	USense_Praca_do_Marques_Noise_2015_11_10	Praça do Marquês- pole1	10-11-2015	19-11-2015	Dados observacionais (amostra)	26,0 MB	852652
260	USense_Praca_do_Marques_Noise_2015_11_20	Praça do Marquês- pole1	20-11-2015	29-11-2015	Dados observacionais (amostra)	25,9 MB	851448
261	USense_Praca_do_Marques_Noise_2015_11_30	Praça do Marquês- pole1	30-11-2015	09-12-2015	Dados observacionais (amostra)	25,9 MB	851724
262	USense_Praca_do_Marques_Noise_2015_12_10	Praça do Marquês- pole1	10-12-2015	12-03-2016	Dados observacionais (amostra)	29,1 MB	953868
263	USense_Praca_do_Marques_Noise_2016_03_13	Praça do Marquês- pole1	13-03-2016	30-03-2016	Dados observacionais (amostra)	22,9 MB	751463
264	USense_Praca_do_Marques_Noise_2016_04_08	Praça do Marquês- pole1	31-03-2016	08-04-2016	Dados observacionais (amostra)	358 KB	11473
265	USense_Praca_do_Marques_O3_2015_05_24	Praça do Marquês- pole1	24-05-2015	01-06-2016	Dados observacionais (amostra)	10,4 MB	314875
266	USense_Praca_do_Marques_Particles1_2015_05_24	Praça do Marquês- pole1	24-05-2015	02-06-2016	Dados observacionais (amostra)	10,0 MB	334898

267	USense_Praca_do_Marques_Particles2_2015_05_24	Praça do Marquês- pole1	24-05-2015	02-06-2016	Dados observacionais (amostra)	10,5 MB	334898
268	USense_Praca_do_Marques_Precipitation_2015_05_24	Praça do Marquês- pole1	24-05-2015	26-05-2015	Dados observacionais (amostra)	83,1 KB	2741
269	USense_Praca_do_Marques_Solar radiation	Praça do Marquês- pole1			Dados observacionais (amostra)	97 bytes	0
270	USense_Praca_do_Marques_Temperature_2015_05_24	Praça do Marquês- pole1	24-05-2015	02-06-2016	Dados observacionais (amostra)	9,34 MB	306519
271	USense_Praca_do_Marques_Wind direction_2015_05_24	Praça do Marquês- pole1	24-05-2015	26-05-2015	Dados observacionais (amostra)	79,0 KB	2664
272	USense_Praca_do_Marques_Wind speed_2015_05_24	Praça do Marquês- pole1	24-05-2015	26-05-2015	Dados observacionais (amostra)	84,5 KB	2790
273	USense_Praca_Velasquez_CO_2015_10_01	Praça Velasquez	01-10-2015	09-05-2016	Dados observacionais (amostra)	853 KB	25047
274	USense_Praca_Velasquez_Humidity_2015_10_01	Praça Velasquez	01-10-2015	03-06-2016	Dados observacionais (amostra)	658 KB	21146
275	USense_Praca_Velasquez_Luminosity_2015_10_01	Praça Velasquez	01-10-2015	09-05-2016	Dados observacionais (amostra)	932 KB	27302
276	USense_Praca_Velasquez_NO2_2015_10_01	Praça Velasquez	01-10-2015	15-04-2016	Dados observacionais (amostra)	851 KB	24994
277	USense_Praca_Velasquez_Noise_2015_10_01	Praça Velasquez	01-10-2015	14-10-2015	Dados observacionais (amostra)	29,1 MB	956732
278	USense_Praca_Velasquez_Noise_2015_10_15	Praça Velasquez	15-10-2015	03-06-2016	Dados observacionais (amostra)	14,6 MB	479489
279	USense_Praca_Velasquez_O3_2015_10_01	Praça Velasquez	01-10-2015	15-04-2016	Dados observacionais (amostra)	851 KB	24973
280	USense_Praca_Velasquez_Particles1_2015_10_01	Praça Velasquez	01-10-2015	03-06-2016	Dados observacionais (amostra)	911 KB	29319
281	USense_Praca_Velasquez_Particles2_2015_10_01	Praça Velasquez	01-10-2015	03-06-2016	Dados observacionais (amostra)	961 KB	29319
282	USense_Praca_Velasquez_Precipitation_2015_10_01	Praça Velasquez	01-10-2015	03-06-2016	Dados observacionais (amostra)	815 KB	26889
283	USense_Praca_Velasquez_Solar radiation_2015_10_01	Praça Velasquez	01-10-2015	09-05-2016	Dados observacionais (amostra)	451 KB	13942
284	USense_Praca_Velasquez_Temperature_2015_10_01	Praça Velasquez	01-10-2015	03-06-2016	Dados observacionais (amostra)	708 KB	22686
285	USense_Praca_Velasquez_Wind direction_2015_10_01	Praça Velasquez	01-10-2015	19-10-2015	Dados observacionais (amostra)	215 KB	7247
286	USense_Praca_Velasquez_Wind speed_2015_10_01	Praça Velasquez	01-10-2015	03-06-2016	Dados observacionais (amostra)	878 KB	29013
287	USense_Reitoria_CO_2015_07_15	Reitoria	15-07-2015	11-12-2015	Dados observacionais (amostra)	6,27 MB	188718
288	USense_Reitoria_Humidity_2015_07_15	Reitoria	15-07-2015	11-12-2015	Dados observacionais (amostra)	5,74 MB	188508
289	USense_Reitoria_Luminosity_2015_07_15	Reitoria	15-07-2015	11-12-2015	Dados observacionais (amostra)	5,98 MB	177192
290	USense_Reitoria_NO2_2015_07_16	Reitoria	16-07-2015	11-12-2015	Dados observacionais (amostra)	3,18 MB	95818
291	USense_Reitoria_Noise	Reitoria			Dados observacionais (amostra)	67 Bytes	0
292	USense_Reitoria_O3_2015_07_16	Reitoria	16-07-2015	11-12-2015	Dados observacionais (amostra)	3,81 MB	113876
293	USense_Reitoria_Particles1_2015_07_15	Reitoria	15-07-2015	11-12-2015	Dados observacionais (amostra)	6,23 MB	198105
294	USense_Reitoria_Particles2_2015_07_15	Reitoria	15-07-2015	11-12-2015	Dados observacionais (amostra)	6,37 MB	198105
295	USense_Reitoria_Precipitation_2015_07_15	Reitoria	15-07-2015	23-10-2015	Dados observacionais (amostra)	3,00 MB	101421
296	USense_Reitoria_Solar radiation_2015_07_15	Reitoria	15-07-2015	23-10-2015	Dados observacionais (amostra)	3,63 MB	107697
297	USense_Reitoria_Temperature_2015_07_15	Reitoria	15-07-2015	11-12-2015	Dados observacionais (amostra)	5,81 MB	190868
298	USense_Reitoria_Wind direction_2015_07_15	Reitoria	15-07-2015	23-10-2015	Dados observacionais (amostra)	2,67 MB	90743
299	USense_Reitoria_Wind speed_2015_07_15	Reitoria	15-07-2015	23-10-2015	Dados observacionais (amostra)	3,99 MB	134548
300	USense_Rua_das_Flores_CO_2015_06_25	Rua das Flores	25-06-2015	24-01-2017	Dados observacionais (amostra)	6,57 MB	197542
301	USense_Rua_das_Flores_Humidity_2014_07_22	Rua das Flores	22-07-2014	24-01-2017	Dados observacionais (amostra)	8,81 MB	288983
302	USense_Rua_das_Flores_Luminosity_2015_06_25	Rua das Flores	25-06-2015	24-01-2017	Dados observacionais (amostra)	6,70 MB	199094
303	USense_Rua_das_Flores_NO2_2015_06_25	Rua das Flores	25-06-2015	24-01-2017	Dados observacionais (amostra)	6,62 MB	198677
304	USense_Rua_das_Flores_Noise_2014_08_01	Rua das Flores	01-08-2014	03-08-2016	Dados observacionais (amostra)	17,8 MB	585203
305	USense_Rua_das_Flores_Noise_2016_09_01	Rua das Flores	01-09-2016	15-11-2016	Dados observacionais (amostra)	20,3 MB	663419

306	USense_Rua_das_Flores_Noise_2016_11_16	Rua das Flores	16-11-2016	24-01-2017	Dados observacionais (amostra)	24,5 MB	800883
307	USense_Rua_das_Flores_O3_2015_06_25	Rua das Flores	25-06-2015	24-01-2017	Dados observacionais (amostra)	6,61 MB	198663
308	USense_Rua_das_Flores_Particles1_2015_06_25	Rua das Flores	25-06-2015	24-01-2017	Dados observacionais (amostra)	6,19 MB	204170
309	USense_Rua_das_Flores_Particles2_2015_06_25	Rua das Flores	25-06-2015	24-01-2017	Dados observacionais (amostra)	6,33 MB	204170
310	USense_Rua_das_Flores_Precipitation_2014_07_22	Rua das Flores	22-07-2014	24-01-2017	Dados observacionais (amostra)	10,1 MB	341574
311	USense_Rua_das_Flores_Solar radiation_2014_07_22	Rua das Flores	22-07-2014	24-01-2017	Dados observacionais (amostra)	12,1 MB	361033
312	USense_Rua_das_Flores_Temperature_2014_07_22	Rua das Flores	22-07-2014	24-01-2017	Dados observacionais (amostra)	8,90 MB	292527
313	USense_Rua_das_Flores_Wind direction_2014_07_25	Rua das Flores	25-07-2014	24-01-2017	Dados observacionais (amostra)	9,18 MB	316847
314	USense_Rua_das_Flores_Wind speed_2014_07_22	Rua das Flores	22-07-2014	24-01-2017	Dados observacionais (amostra)	10,6 MB	359169
315	USense_Trindade_CO_2015_08_03	Trindade	03-08-2015	18-07-2017	Dados observacionais (amostra)	567 KB	16523
316	USense_Trindade_Humidity_2015_08_03	Trindade	03-08-2015	04-01-2017	Dados observacionais (amostra)	598 KB	19360
317	USense_Trindade_Luminosity_2015_08_03	Trindade	03-08-2015	15-03-2016	Dados observacionais (amostra)	330 KB	9690
318	USense_Trindade_NO2_2015_08_03	Trindade	03-08-2015	27-07-2017	Dados observacionais (amostra)	425 KB	12466
319	USense_Trindade_Noise_2015_08_03	Trindade	03-08-2015	19-03-2016	Dados observacionais (amostra)	20,2 MB	662487
320	USense_Trindade_Noise_2016_04_09	Trindade	20-03-2016	09-04-2016	Dados observacionais (amostra)	28,4 MB	930891
321	USense_Trindade_Noise_2016_04_10	Trindade	10-04-2016	29-04-2016	Dados observacionais (amostra)	23,8 MB	782251
322	USense_Trindade_Noise_2016_04_30	Trindade	30-04-2016	19-05-2016	Dados observacionais (amostra)	25,3 MB	830090
323	USense_Trindade_Noise_2016_05_20	Trindade	20-05-2016	08-06-2016	Dados observacionais (amostra)	25,0 MB	819464
324	USense_Trindade_Noise_2016_08_09	Trindade	09-08-2016	18-06-2017	Dados observacionais (amostra)	29,5 MB	966956
325	USense_Trindade_O3_2015_08_03	Trindade	03-08-2015	02-01-2017	Dados observacionais (amostra)	425 KB	12470
326	USense_Trindade_Particles1_2015_08_03	Trindade	03-08-2015	19-06-2017	Dados observacionais (amostra)	847 KB	27695
327	USense_Trindade_Particles2_2015_08_03	Trindade	03-08-2015	19-06-2017	Dados observacionais (amostra)	874 KB	27695
328	USense_Trindade_Temperature_2015_08_03	Trindade	03-08-2015	04-01-2017	Dados observacionais (amostra)	613 KB	19644

Anexo III. Guião de Entrevista aos Investigadores do Projeto UrbanSense

GUIÃO DE ENTREVISTA AOS INVESTIGADORES DO PROJETO URBANSENSE

Investigador-

Unidade a que pertence-

Contacto-

Setor/sensor-

Data-

Inventário

1- **Descrição da área do projeto de investigação de que é responsável.**

2- **Pode fazer uma descrição dos *data set* com o qual está a trabalhar?** ***Características dos dados: tipologia dos dados, quantidade, etc.***

Ciclo de vida dos dados

3- **O que acontece durante o processo de recolha dos dados?**

4- Que instrumentos são utilizados na recolha dos dados?

5- **Quais os fatores que podem influenciar a qualidade dos dados?**

Avaliação dos dados

6- **Quais os dados mais importantes a analisar, disponibilizar e preservar (gerir e manter a longo prazo)? Que parâmetros podem ser utilizados para distinguir dados "válidos" ou "inválidos"?**

Descrição

7- **Que documentação costumam utilizar? Têm alguma estratégia para dar contexto aos dados? Por exemplo utilizam anotações nos dados?**

8- **Que nomenclatura atribuem aos dados? Por exemplo, têm algumas sugestões de nomes para as colunas do ficheiro csv.?**

9- Recomendam ou tencionam fazer alguma normalização dos *data set*?

Visualização

10- **O que recomendam em termos de visualização dos dados?**

Controlo de Acesso

11- **Alguma vez teve alguma experiência com a reutilização de dados noutros projetos?**

12- **Em que projetos de investigação os dados podem ser reutilizados?**

13- **Relativamente à partilha de dados com a população em geral, que impacto acha que estes dados podem ter para a sociedade?**

14- **Quem pode aceder aos dados e quando? A partir de que data podem aceder aos dados?**

Impacto/Resultados

15- **Que mais valia podem obter os utilizadores a partir da utilização dos dados?**

16- **Considera importante obter informação acerca da utilização dos dados, dos utilizadores que utilizaram os dados?**